Universidad Autónoma de Madrid
Facultad de Ciencias
Departamento de Física Teórica

Searches for top-antitop quark resonances in semileptonic final states with the CMS detector

---

Memoria de Tesis Doctoral realizada por

Marino Missiroli

presentada ante el Departamento de Física Teórica
de la Universidad Autónoma de Madrid
para la obtención del Título de Doctor en Ciencias.

Tesis Doctoral dirigida por

Prof. Jorge Fernández de Trocóniz Acha

del Departamento de Física Teórica
de la Universidad Autónoma de Madrid

Madrid, Marzo de 2017

Abstract

This thesis presents the results of two searches for a top-antitop quark ($t\bar{t}$) resonance in semileptonic final states using data collected by the Compact Muon Solenoid (CMS) experiment at the CERN Large Hadron Collider (LHC). Evidence of resonant $t\bar{t}$ production would represent a clear sign of new physics beyond the Standard Model of particle physics.

The first search is based on the full data set recorded by the CMS experiment in proton-proton (pp) collisions at a center-of-mass energy of $\sqrt{s} = 8$ TeV during the LHC Run-1, for a total integrated luminosity of 19.7 fb^{-1}. The second search considers 2.6 fb^{-1} of pp collisions data recorded by the CMS experiment in the first year of the LHC Run-2 (2015) at the higher center-of-mass energy of $\sqrt{s} = 13$ TeV. Both analyses make use of state-of-the-art techniques for the identification of top quarks produced with large transverse momentum. This approach maximizes the sensitivity of the analyses for high-mass $X \to t\bar{t}$ resonances ($M_X \gtrsim 1$ TeV), which are characterized by the production of an increasing fraction of top quarks in the boosted regime.

No significant evidence for resonant $t\bar{t}$ production is found in the data considered and the invariant mass spectrum of the reconstructed $t\bar{t}$ system is used to set upper limits on the production cross section of a $X \to t\bar{t}$ resonance in various new physics models.

Resumen

Esta tesis presenta los resultados de dos búsquedas de una resonancia de quarks top-antitop ($t\bar{t}$) en estados finales semileptónicos utilizando datos recogidos por el experimento Compact Muon Solenoid (CMS) en el Large Hadron Collider (LHC) del CERN. Pruebas de la producción resonante de un par $t\bar{t}$ representarían una clara señal de nueva física más allá del Modelo Estándar de la física de partículas.

La primera búsqueda se basa en la muestra de datos completa registrada por el experimento CMS en colisiones protón-protón (pp) a una energía en el centro de masas de $\sqrt{s} = 8$ TeV durante el Run-1 del LHC, para una luminosidad integrada total de 19.7 fb^{-1}. La segunda búsqueda considera 2.6 fb^{-1} de datos de colisiones pp registrados por el experimento CMS en el primer año del Run-2 del LHC (2015) a la mayor energía en el centro de masas de $\sqrt{s} = 13$ TeV. Ambos análisis utilizan técnicas de vanguardia para la identificación de quarks top producidos con elevado momento transverso. Este enfoque maximiza la sensibilidad de los análisis para resonancias $X \to t\bar{t}$ de alta masa ($M_X \gtrsim 1$ TeV), las cuales se caracterizan por la producción de una fracción creciente de quarks top "boosted".

Las dos búsquedas muestran que no hay pruebas significativas de la producción de una resonancia $t\bar{t}$ en las muestras de datos estudiadas y el espectro de masa invariante del sistema $t\bar{t}$ reconstruido se utiliza para poner límites superiores sobre la sección eficaz de producción de una resonancia $X \to t\bar{t}$ en el marco de varios modelos de nueva física.

Contents

Preface

Our current understanding of the physics of elementary particles is encoded in the so-called Standard Model (SM) of particle physics. The SM is the result of decades of continuous interplay between increasingly precise experimental measurements and the development of the theoretical framework to explain and predict such observations. This model currently provides the framework to understand the properties of all known elementary particles and their interactions under the strong, weak and electromagnetic forces.

The aim to test the predictions of the SM at new energy scales and to search for evidence of new physics motivated the construction of the Large Hadron Collider (LHC) at CERN, the most powerful particle collider ever built. At the time of its startup in 2009, the main goal of the LHC physics program was to search for the Higgs boson, the only elementary particle predicted by the SM that still had not been observed at the time. The discovery in 2012 of a new scalar particle compatible with the SM Higgs boson, based on the analysis of data recorded by the CMS [1] and ATLAS [2] Collaborations during the LHC Run-1, represents the most important milestone of the LHC physics program to date, as it completes the spectrum of the elementary particles of the SM. On the other hand, no clear evidence for physics phenomena beyond the SM (BSM) has been established so far by the LHC experiments. Thus, after the Higgs discovery, the quest for the first evidence of BSM physics assumes an increasingly important role as the LHC program moves forward.

There are reasons suggesting that new physics could be strongly related to the top quark sector. The top quark is the heaviest known elementary particle and it is the only quark whose mass is of the order of the electroweak scale. As a consequence of its unique properties, several theoretical extensions of the SM predict the existence of new massive particles which couple to the top quark and decay preferentially to a top-antitop quark ($t\bar{t}$) pair. Important examples of such BSM scenarios are represented by models that predict new types of fundamental interactions described by new gauge symmetries and BSM theories based on the introduction of additional space-time dimensions.

From the experimental perspective, the search for a $t\bar{t}$ resonance is the most natural, possibly model-independent, approach to probe the existence of new physics connected to the top quark sector. In the absence of interference effects with SM processes, the distinctive signature of such a signal is the presence of a resonant structure in the invariant mass spectrum of the $t\bar{t}$ system. The most sensitive $t\bar{t}$ decay mode for the discovery of a $X \to t\bar{t}$ resonance is the semileptonic (ℓ + jets) channel, thanks to its relatively large branching fraction ($\sim$ 30%) and the clean event signature provided by the presence of one charged lepton (muon or electron) in the final state.

This thesis describes the results of two searches for resonant $t\bar{t}$ production based on data collected by the CMS experiment at the LHC in proton-proton (pp) collisions at two distinct center-of-mass energies. The first analysis uses the entire data set recorded by the CMS experiment at $\sqrt{s} = 8$ TeV during the LHC Run-1 [3]. The second search consists in a similar analysis performed on the data recorded by the CMS experiment in the first year of the LHC Run-2 (2015), exploiting the higher center-of-mass energy of $\sqrt{s} = 13$ TeV of the LHC collisions [4]. Both searches are performed using the $\ell(= e, \mu) + \text{jets}$ channel, in which one of the two W bosons in the final state decays to leptons, while the other decays hadronically. The two analyses make use of dedicated techniques to improve the identification of top quark decays produced with very large Lorentz boosts, i.e. with a transverse momentum largely exceeding the top quark mass. This maximizes the analyses' sensitivity for high-mass $X \to t\bar{t}$ resonances ($M_X \gtrsim 1$ TeV), which are characterized by the production of an increasing fraction of top quarks in the boosted regime.

The present document is structured as follows. Chapter 1 provides a compact overview of the SM of particle physics and it discusses the basic principles of top quark physics at the LHC. The LHC accelerator complex and the CMS detector are described in Chapter 2. In Chapter 3, we discuss the methods employed for the reconstruction and identification of the physics objects, e.g. muons, electrons and jets, used in the analysis of CMS data. Chapter 4 describes a search for $t\bar{t}$ resonances in semileptonic final states with 19.7 fb^{-1} recorded by the CMS experiment in pp collisions at $\sqrt{s} = 8$ TeV. Chapter 5 presents a second search for resonant $t\bar{t}$ production in $\ell + \text{jets}$ events using an integrated luminosity of 2.6 fb^{-1}, recorded by the CMS experiment during the first year of the LHC Run-2 in pp collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Additional studies aimed at future improvements for $X \to t\bar{t}$ searches are discussed in Chapter 6. One of these studies considers the use of angular observables of the $t\bar{t}$ system to improve the analysis sensitivity and gain insight on the properties of a hypothetical $t\bar{t}$ resonance. A second study is concerned with the treatment of interference effects between the SM $t\bar{t}$ background and a $X \to t\bar{t}$ BSM signal; this study, although general in scope, considers one particular $X \to t\bar{t}$ signal, given by a Kaluza-Klein excitation of a gluon in an extradimensional model. Lastly, the conclusions of this thesis are laid out in the final chapter.

Prólogo

Nuestra comprensión actual de la física de las partículas elementales está contenida en el modelo conocido como el Modelo Estándar (ME) de la física de partículas. El ME es el resultado de décadas de continua sinergia entre mediciones experimentales cada vez más precisas y el desarrollo del marco teórico para explicar y predecir tales observaciones. Este modelo proporciona actualmente el marco para entender las propiedades de todas las partículas elementales conocidas y sus interacciones bajo las fuerzas fuerte, débil y electromagnética.

El objetivo de probar las predicciones del ME a nuevas escalas de energía y de buscar evidencia de nueva física motivó la construcción del Gran Colisionador de Hadrones (Large Hadron Collider, LHC) del CERN, el colisionador de partículas más potente jamás realizado. Desde su puesta en marcha en 2009, el objetivo principal del programa de física del LHC fue la búsqueda del bosón de Higgs, la única partícula elemental predicha por el ME que todavía no había sido observada hasta ese momento. El descubrimiento en 2012 de una nueva partícula escalar compatible con el bosón de Higgs del ME, basado en el análisis de los datos registrados por las Colaboraciones CMS [1] y ATLAS [2] durante el Run-1 del LHC, representa el hito más importante en el programa de física del LHC hasta ahora, porque completa el espectro de las partículas elementales del ME. Por otro lado, los experimentos del LHC no han obtenido todavía evidencia clara de ningún fenómeno de física más allá del ME (MME). Por lo tanto, tras el descubrimiento del bosón de Higgs, la búsqueda de la primera evidencia de física MME asume un papel aún más importante en la continuación del programa del LHC.

Existen razones que sugieren que la nueva física podría estar fuertemente relacionada con la física del quark top. El quark top es la partícula elemental conocida con la mayor masa y es el único quark cuya masa es del orden de la escala electrodébil. Como consecuencia de sus propiedades únicas, varias extensiones teóricas del ME predicen la existencia de nuevas partículas masivas que se acoplan al quark top y que se desintegran principalmente en un par de quarks top-antitop ($t\bar{t}$). Ejemplos importantes de este tipo de escenarios MME son los modelos que predicen nuevos tipos de interacciones fundamentales descritas por nuevas simetrías de gauge y teorías MME basadas en la introducción de nuevas dimensiones del espacio-tiempo.

Desde el punto de vista experimental, la búsqueda de una resonancia de un par $t\bar{t}$ es la estrategia más natural y posiblemente general para probar la existencia de nueva física en relación con el sector del quark top. En ausencia de efectos de interferencia con procesos del ME, la marca distintiva de una tal señal es la presencia de una estructura resonante en el espectro de masa invariante del sistema $t\bar{t}$. El canal de desintegración más sensible para el descubrimiento de una resonancia $X \to t\bar{t}$ es el canal semileptónico (ℓ + jets), gracias a su fracción de desintegración relativamente elevada ($\sim 30\%$) y a la clara signatura experimental dada por la presencia de un leptón cargado (muón o electrón) en el estado final.

Esta tesis describe los resultados de dos búsquedas de producción resonante de un par $t\bar{t}$ basadas en datos registrados por el experimento CMS del LHC en colisiones protón-protón (pp), a dos energías distintas en el centro de masas. El primer análisis utiliza la muestra de datos completa recogida por el experimento CMS a $\sqrt{s} = 8$ TeV durante el Run-1 del LHC [3]. La segunda búsqueda consiste en un análisis similar realizado con los datos registrados por el experimento CMS en el primer año del Run-2 del LHC (2015), sacando partido de la mayor energía en el centro de masas de $\sqrt{s} = 13$ TeV de las colisiones del LHC [4]. Ambas búsquedas se han realizado utilizando el canal $\ell(= e, \mu) + \text{jets}$, en el que uno de los dos bosones W en el estado final se desintegra en leptones, mientras el otro se desintegra en su canal hadrónico. Los dos análisis hacen uso de técnicas especializadas para mejorar la identificación de quarks top producidos con un boost de Lorentz muy elevado, es decir con un momento transverso muy superior a la masa del quark top. Esto maximiza la sensibilidad de los análisis para resonancias $X \to t\bar{t}$ de alta masa ($M_X \gtrsim 1$ TeV), que se caracterizan por la producción de una creciente fracción de quarks top de altas energías.

El presente documento se estructura como sigue. El Capítulo 1 ofrece un resumen del ME de la física de partículas y discute los principios básicos de la física del quark top en el LHC. El complejo del acelerador LHC y el detector CMS se describen en el Capítulo 2. En el Capítulo 3, se discuten los métodos empleados para la reconstrucción y identificación de los objetos físicos, por ejemplo muones, electrones y jets, utilizados en el análisis de los datos de CMS. El Capítulo 4 describe una búsqueda de resonancias $t\bar{t}$ en estados finales semileptónicos con 19.7 fb^{-1} registrados por el experimento CMS en colisiones pp a $\sqrt{s} = 8$ TeV. El Capítulo 5 presenta una segunda búsqueda de producción resonante de un par $t\bar{t}$ en sucesos de tipo $\ell + \text{jets}$ utilizando una luminosidad integrada de 2.6 fb^{-1}, registrada por el experimento CMS durante el primer año del Run-2 del LHC en colisiones pp a una energía en el centro de masas de $\sqrt{s} = 13$ TeV. Estudios adicionales dirigidos a futuras mejoras en búsquedas en el canal $X \to t\bar{t}$ se han estudiado en el Capítulo 6. Uno de estos estudios considera el uso de observables angulares del sistema $t\bar{t}$ para mejorar la sensibilidad del análisis y obtener más información sobre las propiedades de una hipotética resonancia $t\bar{t}$. Un segundo estudio se refiere al tratamiento de los efectos de interferencia entre el fondo $t\bar{t}$ del ME y una señal $X \to t\bar{t}$ de física MME; este estudio, aunque general en alcance, considera una específica señal $X \to t\bar{t}$, correspondiente a una excitación de Kaluza-Klein de un gluón en un modelo extradimensional. Por último, las conclusiones de esta tesis se presentan en el capítulo final.

Chapter 1

Introduction

This first chapter introduces the basics of the Standard Model (SM) of particle physics and it outlines the physics motivation for the experimental work described in this thesis. We first give a compact description of the physics of the SM, followed by an overview of some of the possible indications of the existence of new physics beyond it. We then briefly review some of the most widely known extensions of the SM that predict the existence of a $X \to t\bar{t}$ resonance. The second part of the chapter serves as an introduction to the basics of top quark physics in hadron collider experiments.

1.1 The Standard Model of particle physics

The Standard Model of particle physics describes the properties of all known elementary particles and the dynamics of the strong, weak and electromagnetic interactions. The gravitational force, which is not included in the current formulation of the SM, is the weakest of the four known fundamental forces of Nature and it is completely negligible at the scales explored by current particle physics experiments. The SM is formulated as a quantum field theory based on the principle of gauge symmetry and it is founded on an overwhelming amount of experimental measurements collected during the course of the last decades; for a complete review of the experimental confirmations of the SM we refer to Ref. [5] and references therein. In the following, we briefly summarize the basic structure and properties of the SM; a comprehensive description of the model can be found in modern reviews and introductory textbooks, e.g Ref. [6, 7, 8].

Each of the three fundamental interactions described by the SM (strong, weak and electromagnetic) is associated to a gauge symmetry, as detailed below.

- The strong force is the interaction that acts on the elementary constituents of hadronic matter, i.e. quarks and gluons, and confines them into bound states such as protons and neutrons. This same force is also responsible for the binding of protons and neutrons in the nuclei of atoms. The symmetry group describing this force is the SU(3) group and the conserved charge associated to this symmetry is known as *color charge*.

- The weak interaction is the force responsible for radioactive β decay, the process by which the nucleus of an unstable atom can decay and emit an electron (or positron) and a neutrino. It is described in the SM by a gauge symmetry based on the SU(2) group.

- The electromagnetic (EM) interaction is the force responsible for the binding of electrons to atomic nuclei and it manifests itself, for example, in the interactions between

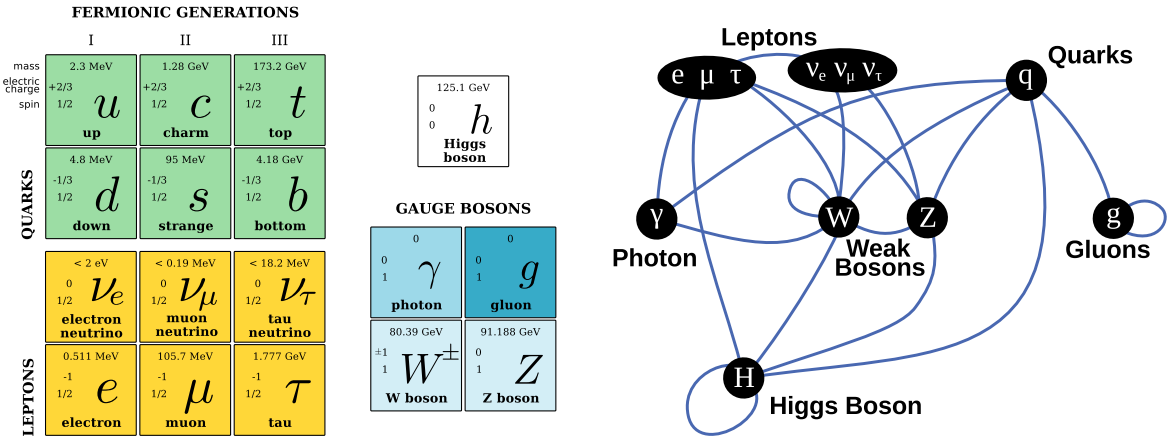


Figure 1.1: Graphic representation of the particle content of the Standard Model (left) and the structure of the particles' interactions (right).

atoms and molecules of ordinary matter. This force is associated to the symmetry group U(1) and the quantity conserved in EM interactions is known as *electric charge*.

The weak and EM interactions arise in the SM as a result of the mechanism of spontaneous breaking of the electroweak (EW) symmetry SU(2) × U(1), mediated by the Higgs field.

The elementary particles of the SM include

- fermions with spin 1/2, which include the basic constituents of ordinary matter,
- vector (spin 1) bosons, which act as mediators of the strong and EW forces, and
- one scalar (spin 0) particle, referred to as the Higgs boson.

Figure 1.1 shows a graphic representation of the particle content of the SM, together with some of the particles' properties, e.g. mass and spin.

The elementary fermions in the SM can be divided in two classes: quarks and leptons. Six different types (flavors) of quarks are present in the SM. These are commonly referred to as up (u), down (d), strange (s), charm (c), bottom (b) and top (t). Quarks are subject to both the strong and electroweak interactions. The (u, c, t) quarks have an electric charge equal to $+2/3\,e$, while the (d, s, b) quarks have a charge of $-1/3\,e$, where e is the elementary electric charge ($e = 1.6 \cdot 10^{-19}$ C)[1]. Similarly to quarks, the SM contains six different types of leptons. There are three types of leptons with electric charge: the electron (e), the muon (μ) and the tau lepton (τ). Each of these three particles is paired in the SM with an electrically neutral lepton, called neutrino (ν_e, ν_μ, ν_τ). Leptons are subject to the weak force, but not to the strong interaction; electrically charged leptons (e, μ, τ) also interact via the electromagnetic force.

Quarks and leptons can be accommodated into three generations, each containing an up-type quark, a down-type quark, a charged lepton and a neutrino, as indicated in Figure 1.1. These three generations are identical in terms of gauge interactions, but they exhibit a large mass hierarchy for both quarks and electrically charged leptons.

The particle content of the SM is completed by the gauge bosons and the Higgs particle. Gauge bosons are spin-1 fields associated to the local symmetries of the SM Lagrangian and they act as mediators of the fundamental forces described by the SM. Eight gluons (g)

[1]Natural units ($c = \hbar = 1$) are used throughout this thesis, unless otherwise specified.

act as mediators of the strong interaction. The mediator associated to the electromagnetic interaction is the photon (γ), while the weak force is mediated by three bosonic particles denoted as $W^{\pm}$ and Z. The gluons and the photon are massless, whereas the $W^{\pm}$ and the Z bosons are massive particles. Finally, the SM contains one massive scalar particle, the Higgs boson. The latter plays a key role in the mechanism through which other particles (specifically, quarks, charged leptons and the gauge bosons of the weak interaction) acquire mass in the SM.

1.1.1 Strong interaction

The strong interaction is the fundamental force which binds quarks together into hadronic particles, such as protons and neutrons. The theory of strong interactions is known as *Quantum Chromodynamics* (QCD) [8].

In the SM, this interaction is implemented via the SU(3) gauge symmetry and the QCD term of the SM Lagrangian can be written as[2]

$$\mathcal{L}_{\mathrm{QCD}} = \sum_{q} \bar{\psi}_q \, \gamma^{\mu} \left(i\partial_{\mu} - \frac{g_{\mathrm{s}}}{2} \, \lambda_a G^a_{\mu} \right) \psi_q - \frac{1}{4} \, G^{a\mu\nu} \, G^a_{\mu\nu} \tag{1.1}$$

where ψ_q corresponds to a SU(3) triplet, which contains three color states for each quark flavor (q); the symbols γ^{μ} and λ_a denote the Dirac and Gell-Mann matrices, respectively; G^a_{μ} corresponds to the eight massless gauge fields of the SU(3) symmetry, the gluons. The field tensor $G^a_{\mu\nu}$ is defined as

$$G^a_{\mu\nu} = \partial_{\mu} \, G^a_{\nu} - \partial_{\nu} \, G^a_{\mu} - g_{\mathrm{s}} \, f_{abc} \, G^b_{\mu} \, G^c_{\nu} \tag{1.2}$$

where the coefficient f_{abc} denotes the structure constants of the SU(3) group. The last term in Equation (1.2) shows that gluons, in addition to mediating the strong interaction between quarks, are self-interacting particles in QCD.

The g_{s} coefficient in Equation (1.1) is the QCD coupling constant, which determines the strength of the strong interaction. This quantity is commonly recast in the more convenient parameter $\alpha_{\mathrm{s}} = g_{\mathrm{s}}^2/(4\pi)$. As a consequence of the renormalization procedure in quantum field theory, the strong coupling constant can be expressed as a quantity depending on the energy scale Q of a given physical process. In the one-loop approximation, this dependence is given by

$$\alpha_{\mathrm{s}}(Q^2) = \frac{1}{b_0 \, \ln(Q^2/\Lambda^2_{\mathrm{QCD}})} \;, \quad \text{with } \; b_0 = \frac{33 - 2n_f}{12\pi} \tag{1.3}$$

where n_f is the number of quark flavors with mass $m_q \ll Q$. The parameter Λ_{QCD} identifies the energy scale at which the QCD coupling constant is expected to become large and, thus, QCD processes cannot be described with fixed-order calculations in the perturbative expansion. The value of Λ_{QCD}, determined indirectly from measurements of $\alpha_{\mathrm{s}}(Q^2)$, ranges between 200 MeV and 350 MeV (depending on the value of n_f). The world average value of $\alpha_{\mathrm{s}}(M_{\mathrm{Z}}^2)$, as determined in Ref. [5], corresponds to $\alpha_{\mathrm{s}}(M_{\mathrm{Z}}^2) = 0.1181 \pm 0.0011$.

Based on Equation (1.3), the value of the strong coupling constant decreases as the energy transfer Q in a given physical process becomes larger. A consequence of this is that colored

[2]The Lagrangian in Equation (1.1) does not include the mass terms for the quark fields, which are a priori forbidden by the gauge principle and arise in the SM as a consequence of the Higgs mechanism, discussed later on in this chapter. For simplicity, we also did not include in $\mathcal{L}_{\mathrm{QCD}}$ the gauge-fixing terms and the CP-violating term of the QCD Lagrangian.

particles behave as weakly interacting particles in high-energy QCD processes ($Q \gg 1$ GeV). This property of the strong interaction, known as *asymptotic freedom*, permits the use of perturbative QCD to determine fixed-order predictions of relevant quantities in high-energy physics experiments, e.g. hard-scattering cross sections.

On the other hand, for sufficiently high values of the strong coupling constant, i.e. at low-energy scales, the strong interaction between quarks and gluons leads to the mechanism of *color confinement*. Experimental observations have confirmed that particles with color charge (quarks and gluons) cannot be isolated and, when they are produced in high-energy collisions, their decay leads to the formation of multiple quark bound states without color charge, called hadrons. This non-perturbative process is known as *hadronization* and it can be simulated in modern event generators for collider physics through the use of phenomenological models.

1.1.2 Electroweak interaction and the Higgs mechanism

The EW interaction is described in the SM by the gauge group SU(2) × U(1). Fermionic fields with left-handed and right-handed chirality are set to transform differently under the electroweak symmetry: left-handed fermions transform as doublets of SU(2), while right-handed fermions are singlets of the same group. This leads to the following set of fields

$$\begin{pmatrix} \nu_L \\ \ell_L \end{pmatrix} , \quad \begin{pmatrix} U_L \\ D_L \end{pmatrix} , \quad \ell_R , \quad U_R , \quad D_R$$

with $\ell \in (e, \mu, \tau)$, $U \in (u, c, t)$ and $D \in (d, s, b)$. Right-handed neutrinos are not included in the SM.

Given this particle content, the unbroken EW Lagrangian derives directly from the principle of gauge invariance under the group SU(2) × U(1), which leads to the introduction of four massless vector fields. The requirement of gauge invariance forbids the introduction of mass terms for the fermionic fields and, also, for the gauge fields of the weak interaction. The mechanism through which these fields acquire mass in the SM is the so-called spontaneous breaking of the electroweak symmetry (electroweak symmetry breaking, EWSB). This procedure requires the introduction of a complex SU(2) doublet Φ, known as the Higgs field, which contains a total of four additional degrees of freedom. The Lagrangian of the Higgs field is given by [5]

$$\mathcal{L}_\Phi = D_\mu \Phi^\dagger D^\mu \Phi - V(\Phi) , \quad \text{with} \quad V(\Phi) = \mu^2 \, \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi \right)^2 . \tag{1.4}$$

where D_μ corresponds to the covariant derivative of the Higgs field under the SU(2) × U(1) group. Gauge invariance also allows for the introduction of the so-called Yukawa terms, which describe the interaction between the Higgs field and fermions and complete the EW sector of the SM.

The properties of the Higgs potential $V(\Phi)$ depend on the values of λ and μ^2. The parameter λ has to be a positive value, in order for the potential to be bounded from below. In the case $\mu^2 > 0$, the potential has a unique minimum at $|\Phi| = 0$, which leaves the EW symmetry intact; conversely, if $\mu^2 < 0$, the Higgs potential is characterized by an infinite set of vacuum states, defined by the condition $\left\langle \Phi^\dagger \Phi \right\rangle = \frac{-\mu^2}{2\lambda}$. This leads to a non-zero vacuum expectation value (VEV) for the Higgs field given by $v/\sqrt{2}$, with $v = \sqrt{-\mu^2/\lambda}$.

The case $\mu^2 < 0$ is the one that applies to the SM, leading to the mechanism of electroweak symmetry breaking; based on experimental measurements, the value of the Higgs field VEV

corresponds to $v \simeq 246$ GeV [5]. The SM Lagrangian remains in general invariant under the gauge group of electroweak interactions, but the vacuum state described by the non-zero VEV of the Higgs field is only invariant under a U(1) gauge symmetry, which ultimately describes the electromagnetic force.

In its vacuum state, the Higgs field can be reexpressed in terms of a perturbation around its non-zero vacuum expectation value (in unitary gauge):

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.5}$$

where the field h is one of the four original degrees of freedom of the Higgs field and, according to Equation (1.4), it corresponds to a physical state, the Higgs particle, a scalar boson with a mass equal to $m_h = v\sqrt{2\lambda}$ and no electric charge. The three remaining degrees of freedom of Φ are absorbed by the gauge bosons of the electroweak interaction. With an appropriate redefinition of the four original EW gauge fields, it is possible to see that the non-zero VEV of the Higgs fields leads to the introduction of mass terms for three bosonic fields which corresponds to the $W^{\pm}$ and Z gauge bosons of the weak interaction; the fourth physical state corresponds to the photon field, which remains massless as the corresponding U(1) gauge symmetry is preserved. Moreover, the introduction of the VEV in the Yukawa coupling of the Higgs to fermions produces mass terms for quarks and electrically charged leptons (e, μ, τ), with the mass of each fermion being proportional to the its coupling to the Higgs boson; neutrinos remain massless in this mechanism because of the absence (in the SM) of their right-handed component.

In the quark sector, mass eigenstates do not correspond to flavor eigenstates, but they are a superposition of the latter. In the basis of mass eigenstates, this implies the existence of flavor-changing charged currents, already at the first order of the perturbative expansion. These transitions between quarks of different flavor are mediated by the electrically charged $W^{\pm}$ gauge bosons. The relative strength of different flavor-changing transitions in the quark sector is encoded in the Cabibbo–Kobayashi–Maskawa (CKM) matrix, V_{CKM}, defined as

$$\begin{pmatrix} d' \\ s' \\ b \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad \text{with} \quad V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.6}$$

where (d', s', b') denotes the down-type quark flavor eigenstates and (d, s, b) corresponds to the mass eigenstates of down-type quarks. The probability of a quark of flavor i to decay to a quark of flavor j is proportional to $|V_{ij}|^2$. The elements of the CKM matrix can be parameterized by four independent parameters that can only be determined experimentally. The measured magnitudes of the nine CKM matrix elements, obtained using a global fit of several independent measurements and imposing the unitarity of the V_{CKM} matrix [5], correspond to

$$V_{\text{CKM}} = \begin{pmatrix} 0.97434^{+0.00011}_{-0.00012} & 0.22506 \pm 0.00050 & 0.00357 \pm 0.00015 \\ 0.22492 \pm 0.00050 & 0.97351 \pm 0.00013 & 0.0411 \pm 0.0013 \\ 0.00875^{+0.00032}_{-0.00033} & 0.0403 \pm 0.0013 & 0.99915 \pm 0.00005 \end{pmatrix} \tag{1.7}$$

These values reveal the strong hierarchy of flavor-changing charged currents, in favor of transitions between quarks of the same generation.

Flavor mixing in the lepton sector is absent (at leading order) if neutrinos are set to be massless.

The Lagrangian of the EW sector after EWSB describes the structure of the fundamental interactions involving fermions, electroweak gauge bosons and the Higgs boson. The main properties of these interactions can be summarized as follows.

- The electrically-charged $\mathrm{W}^{\pm}$ gauge bosons interact with every generation of quarks and leptons, but they only couple to left-handed fermions. This interaction is described in the SM Lagrangian by terms of the form

$$\frac{g}{\sqrt{2}} \sum_{u,d} \overline{u_L}\, \gamma^{\mu}\, \mathrm{W}^{+}_{\mu}\, V_{ud}\, d_L + \text{h. c.} \quad \text{and} \quad \frac{g}{\sqrt{2}} \sum_{\ell} \overline{\nu_{\ell,L}}\, \gamma^{\mu}\, \mathrm{W}^{+}_{\mu}\, \ell_L + \text{h. c.}\,, \tag{1.8}$$

  where g corresponds to the coupling constant of the SU(2) gauge group. The sum in the first term runs over the three flavors of up-type (u) and down-type (d) quarks and V_{ud} denotes the corresponding element of the CKM matrix; similarly, the second term runs over the three generations of leptons (ℓ). The interactions mediated by the $\mathrm{W}^{\pm}$ bosons are the only ones allowing flavor-changing processes at tree level in the quark sector.

- The Z boson interacts with every type of fermion as well, but its effective couplings to left-handed and right-handed fermions are different. This interaction is described by in the SM Lagrangian by a term proportional to

$$\bar{\psi}_f\, \gamma^{\mu}\, \mathrm{Z}_{\mu}\, \left(g_V^f + \gamma^5 g_A^f \right)\, \psi_f \tag{1.9}$$

  where $g_V^f = (g_R^f + g_L^f)/2$ and $g_A^f = (g_R^f - g_L^f)/2$. The symbols g_R^f and g_L^f denote the Z boson couplings to the left-handed and right-handed components of fermion f, respectively, and their values depend on the type of fermion.

- The photon couples to fermions with non-zero electric charge.

- The Higgs boson couples to fermions with non-zero mass and each Higgs-fermion coupling is proportional to the value of the fermion mass.

- The EW gauge bosons and the Higgs boson couple with each other via interactions involving three bosons (triple gauge interactions) and four bosons (quartic gauge interactions). The Higgs boson is also a self-interacting particle, as its potential contains triple and quartic self-interaction terms.

1.2 Open questions on physics beyond the SM

The SM provides a consistent framework to describe the overwhelming majority of the experimental results collected in the field of particle physics to date. On the other hand, there are theoretical arguments and some experimental evidence suggesting that the SM may be only part of a more general physics model, which is yet to be unveiled.

The hierarchy problem and fine-tuning in the Higgs sector

One of the main theoretical limits attributed to the SM is related to the Higgs sector and it is known as the "hierarchy (or naturalness) problem" [9, 10, 11]. The issue lies in the fact that, in the presence of new physics coupled to the Higgs field and characterized by an energy scale Λ_{NP}, the Higgs mass would receive radiative corrections of the order of Λ_{NP}. If the Λ_{NP}

scale were much higher than the scale of EW interactions, this would require the presence of an unnaturally high level of fine-tuning in the cancellation of these quantum corrections, in order to guarantee the relatively low value of the Higgs mass, which is of the order of 10^2 GeV. An example for a possible value of $\Lambda_{\rm NP}$ is the characteristic scale of quantum gravity, i.e. the Planck scale, which is extremely higher compared to the Higgs mass. From a less technical perspective, this theoretical puzzle reflects the lack of a deep understanding of why the measured mass of the Higgs boson is as small as it is. The "hierarchy problem" could be regarded as one of the theoretical arguments to expect the emergence of new physics at the TeV scale, an energy scale much closer to the Higgs mass and the scale of EW interactions.

Dark Matter

Cosmological observations suggest that ordinary (baryonic) matter constitutes only a small fraction of the total amount of matter in the Universe. The majority of such matter, whose presence is inferred from its gravitational effects, is not subject to electromagnetic interactions and, for this reason, it is referred to as *Dark Matter* (DM) [12]. It remains to be assessed whether or not an explanation for this unidentified type of matter can be found in the context of particle physics and, in that case, if the presence of DM is linked to the existence of new elementary particles not predicted by the SM.

Flavor mixing in the lepton sector and neutrino masses

The phenomenon of neutrino oscillations has been firmly established by experiments in recent years (see the relevant section in Ref. [5] and references therein). One of the implications of neutrino mixing is that at least two of the three SM neutrinos have a non-zero mass and neutrino masses differ across their three generations. Although the SM framework can in principle accommodate for non-zero neutrino masses by means of minimal generalizations, the mechanism behind the generation of neutrino masses remains to be determined. This aspect is closely related to that of clarifying whether neutrinos are Dirac fermions, like all the other fermions of the SM, or Majorana particles, i.e. they correspond to their own antiparticle. The results obtained by dedicated experiments are still not conclusive in this respect.

Free parameters in the SM

A theoretical limit of the SM is related to the presence of a large number of free parameters (19), which can only be determined experimentally. These are given by the twelve masses of quarks and charged leptons, the four coefficients parameterizing the CKM matrix, two parameters in the QCD Lagrangian (the coupling constant $g_{\rm s}$ and the coefficient $\theta_{\rm QCD}$ of the CP-violating term in QCD), and a total of four parameters in the EW and Higgs sectors. The SM does not provide an explanation for the measured values of these parameters. Moreover, the numerical values of some of them seem to suggests the presence of additional symmetries or mechanisms which are not explicitly described by the SM. Two examples in this context are the hierarchy between the masses of the three generations of fermions and the very small value measured for the coefficient $\theta_{\rm QCD}$ associated to the CP-violating term in QCD ($\theta_{\rm QCD} < 10^{-9}$); the non-observation of CP violation in the strong sector and, thus, the apparent fine-tuning of $\theta_{\rm QCD}$ represents a theoretical puzzle known as the "strong CP problem".

Gravitational force

The SM does not describe one of the four known fundamental interactions in Nature, the gravitational force. On the one hand, the strength of gravitational effects in particles' interactions would be comparable to that of the other fundamental forces only at the so-called Planck scale ($m_{\mathrm{P}} \simeq 10^{19}$ GeV), so these effects can be completely neglected in current particle physics experiments. On the other hand, a well-tested model that includes a description of the gravitational interaction in the framework of particle physics, if at all possible, remains to be determined.

1.3 Extensions of the SM predicting $X \to t\bar{t}$ resonances

Several extensions of the SM have been developed in the last decades to address some of the potential deficiencies of the SM. These models, referred to as theories *beyond the Standard Model* (BSM), predict new physics phenomena and, in most cases, the existence of new elementary particles.

This section presents an overview of some of the most widely known BSM scenarios that predict the existence of new physical states decaying to a $t\bar{t}$ pair, these being the most relevant BSM models for the experimental searches presented in this thesis. The main phenomenological consequences expected for a $X \to t\bar{t}$ signal in these BSM scenarios are also discussed.

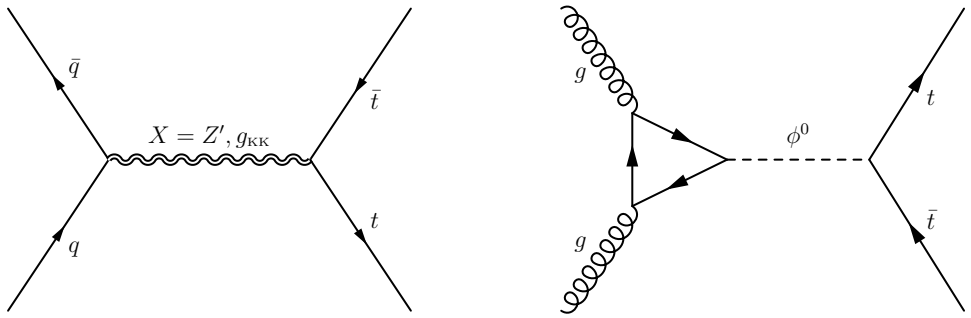
- **Models with extended gauge symmetries**

  A multitude of BSM scenarios rely on the introduction of new gauge symmetries in order to elucidate some of the properties of the SM. Examples of such models are Grand Unified Theories [13,14,15] and models in which a new strong interaction is involved in the breaking of the EW symmetry, e.g. technicolor models [16,17]. Common to all these extensions of the SM is their prediction of a new heavy neutral gauge boson, generally referred to as a Z' boson, which has been the object of extensive phenomenological studies [18, 19]. In many of these BSM scenarios the new boson can decay to light fermions, e.g. lepton pairs, but there are also several models in which the Z' couplings to light fermions are expected to be small and this new particle can preferentially decay to a $t\bar{t}$ pair. A relevant example in this context is the leptophobic Z' boson of the *topcolor* model [20,21], which has been often used as a benchmark in past searches for $t\bar{t}$ resonances at colliders (see Section 1.5).

  The Z' boson in these models is a color-singlet with spin 1. At the LHC, this type of particle can be produced in the s-channel via $q\bar{q}$ annihilation, as shown in Figure 1.2(a). The mass and decay width of the Z' boson are generally free parameters of the model. Since the Z' is a color-singlet, the production of a $\mathrm{Z}' \to t\bar{t}$ signal does not interfere with the SM $t\bar{t}$ background at leading order. Therefore, this BSM signal would appear as a resonance in the $M_{t\bar{t}}$ spectrum on top of the smoothly-falling distribution expected from SM processes.

- **Extra-dimensional models**

  Several extensions of the SM are based on the introduction of additional space-time dimensions (or extra-dimensions). These models provide a natural explanation for the weakness of the gravitational force.



(a) (b)

Figure 1.2: Examples of the dominant processes for the production of a massive BSM resonance decaying to a $t\bar{t}$ pair in a hadron collider experiment: (a) spin-1 particle (Z′ boson or KK gluon, g_{KK}) produced via quark-antiquark annihilation, (b) spin-0 (Higgs-like) particle produced via gluon-fusion.

One of the most important examples of such models is the Randall-Sundrum (RS) model [22, 23]. This model predicts the existence of many new massive states, such as Kaluza-Klein (KK) excitations of SM particles and gravitons.

In the most common parameterization of the RS model, the KK excitation of the SM gluon, or KK gluon (g_{KK}), has an enhanced coupling to the top quark and it preferentially decays to a $t\bar{t}$ pair. The cross sections for the production of other BSM particles decaying to a $t\bar{t}$ pair, e.g. KK gravitons, are expected to be much smaller compared to that for the production of a KK gluon resonance. The latter is thus the principal signal to test the RS model in the $t\bar{t}$ channel [24]. The production mechanism for a massive KK gluon at the LHC is dominated by quark-antiquark annihilation (see Figure 1.2). The mass of a KK gluon ($M_{g_{\mathrm{KK}}}$) is a free parameter of the model and its decay width, which depends on the value of $M_{g_{\mathrm{KK}}}$, varies between 15% and 20% of the KK gluon mass. Since KK gluons are colored particles, this type of signal interferes with the SM background $q\bar{q} \to t\bar{t}$ already at leading order. Because of this interference term, the experimental signature of a $g_{\mathrm{KK}} \to t\bar{t}$ signal potentially differs from that of a color-singlet resonance. On the other hand, a dedicated study described in Chapter 6 will show that neglecting the interference term for the $g_{\mathrm{KK}} \to t\bar{t}$ signal is acceptable in first approximation.

- **Two-Higgs-Doublet models**

  New (pseudo)scalar states coupled to the top quark are predicted by a class of models in which the Higgs sector is extended to include a second Higgs doublet. These models, generally referred to as *Two-Higgs-Doublet* (2HDM) models [25], predict the existence of five spin-0 particles. These correspond to a state compatible with the Higgs boson of the SM, two scalar particles with electric charge ($H^{\pm}$), a CP-even neutral scalar (a heavier Higgs-like particle, H) and a CP-odd neutral pseudoscalar (A). For certain parameterizations of this extended Higgs sector, one of these new massive neutral particles may preferentially decay to a $t\bar{t}$ pair [26].

  In general, a massive Higgs-like particle (ϕ^0) decaying to a $t\bar{t}$ pair would be predom-

inantly produced at the LHC via gluon-fusion in the process $gg \to \phi^0 \to t\bar{t}$ shown in Figure 1.2(b). In this scenario, there is a large interference between the $\phi^0 \to t\bar{t}$ signal and the SM $t\bar{t}$ background, which is also produced mainly via gluon-fusion at the LHC. The presence of this interference term leads to a peak-dip structure in the invariant mass distribution of the $t\bar{t}$ system, around the nominal value of the resonance mass. Because of the large impact of interference effects expected in this BSM scenario, the search for this type of signal differs substantially from a search for a local excess in the $M_{t\bar{t}}$ distribution. For this reason, searches for heavy Higgs-like $t\bar{t}$ resonances are carried out by the CMS and ATLAS [27] Collaborations in dedicated analyses, which are beyond the scope of this thesis.

This overview shows that the characteristic signature of a $X \to t\bar{t}$ signal in a variety of BSM scenarios, with the exception of a heavy Higgs-like $\phi^0 \to t\bar{t}$ resonance, is given by the presence of a resonant structure in the invariant mass distribution of the $t\bar{t}$ system [28]. Regardless of the specifics of the model, the two most important parameters defining the kinematics of this signal are the mass and the decay width of the hypothetical new particle.

This motivates the experimental searches for a resonance in the $M_{t\bar{t}}$ spectrum presented in Chapter 4 and Chapter 5. The impact of interference effects in these searches are quantified in a dedicated study described in Chapter 6.

1.4 Top quark physics at hadron colliders

The top quark, the up-type quark of the third generation of fermions in the SM, was discovered by the CDF [29] and D0 [30] Collaborations at the Tevatron collider in 1995. It is the heaviest elementary particle known to date; the current most precise determination for the value of its mass, which is not predicted by the SM, is $m_t = 173.34 \pm 0.27\,(\text{stat}) \pm 0.71\,(\text{syst})$, based on the combination of measurements from the Tevatron and LHC experiments [31].

In the following, we review the basic properties of the top quark and its phenomenology in hadron collider experiments, from its production mechanisms to the signatures originating from its decay. A particular focus is given to the physics of $t\bar{t}$ events in the semileptonic final state, which is the channel used in the experimental analyses described in this work.

Hard-scattering cross section and Parton Distribution Functions

Before discussing the specifics of top quark physics, we briefly introduce some of the general tools used to calculate the cross section of a given process in a hadron collider experiment.

The inclusive cross section for the production of a generic final state F in the collision of two hadrons A and B, i.e. $A + B \to F + X$, can be calculated according to the following expression

$$\sigma_{AB\to F+X}(s) = \sum_{a,b} \int dx_1\, dx_2\; f_a^{(A)}(x_1, \mu_F^2)\; f_b^{(B)}(x_2, \mu_F^2)\; \hat{\sigma}_{ab\to F}\left(x_1 x_2 s,\, \mu_R^2,\, \mu_F^2\right) \qquad (1.10)$$

where $\sqrt{s}$ is the c. o. m. energy of the hadronic collision. Equation (1.10) describes the factorization between the hard-scattering process, described by the parton-level cross section $\hat{\sigma}$, and the dynamics of the constituents of the initial-state hadrons, which is parameterized by the Parton Distribution Functions (PDF) $f_i^{(H)}(x, \mu)$.

The PDF $f_i^{(H)}(x, \mu)$ can be interpreted as the probability of finding a parton i ($q, \bar{q}, g$ with $q = u, d, s, c, b$) carrying a fraction x of the momentum of its mother hadron H, in a

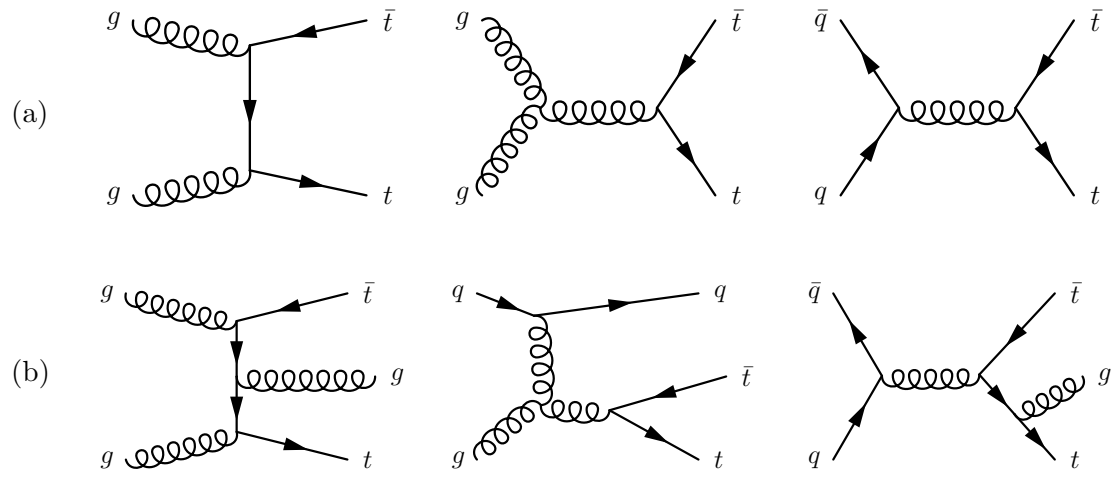


Figure 1.3: Feynman diagrams for the production of a $t\bar{t}$ pair in a hadron-hadron collision: (a) leading-order diagrams, (b) three examples of NLO diagrams in QCD.

process with momentum transfer μ. The numerical values of the PDFs can be determined from experimental data in global fits based on the combination of multiple independent measurements. These measurements show that for small values of Q^2 a large fraction of the hadron momentum is carried by its valence quarks, whereas at high energies most of the hadron momentum is carried by the so-called *sea partons*, i.e. gluons and virtual $q\bar{q}$ pairs.

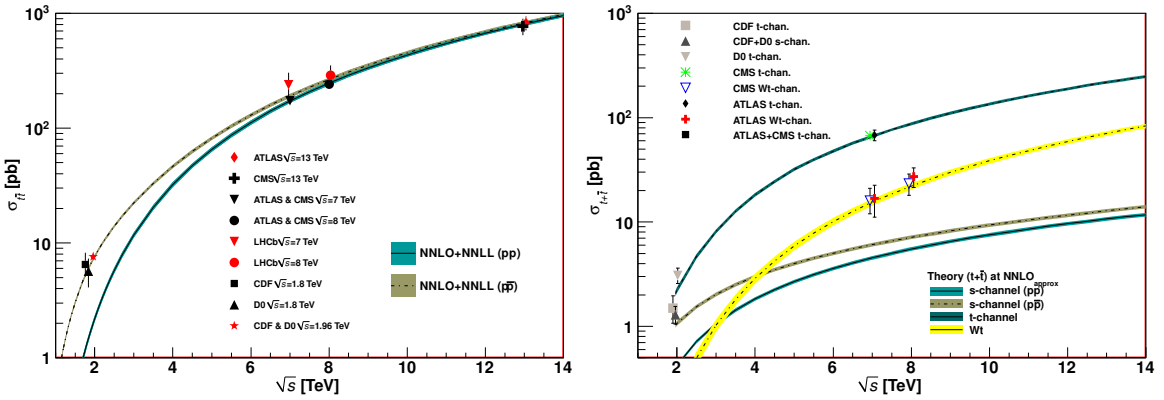
The hard-scattering cross section $\hat{\sigma}$ describes the parton-level process $ab \to F$ and it can be calculated analytically via the use of perturbation theory and resummation techniques. The cross section depends on two arbitrary energy scales, μ_F and μ_R, referred to in the following as the *factorization scale* and *renormalization scale*, respectively. This dependence is an artifact of fixed-order calculations and it introduces a theoretical uncertainty in the determination of $\hat{\sigma}_{ab\to F}$. A method to reduce the size of this uncertainty is to set these scales to the typical energy scale of the hard-scattering process. The corresponding uncertainty is estimated by varying the two scales up and down, usually by a factor of 1/2 and 2. In general, the size of this uncertainty decreases as higher-order corrections are included in the calculation of $\hat{\sigma}_{ab\to F}$.

1.4.1 Production mechanisms

The two main processes responsible for the production of top quarks in a hadron-hadron collision are $t\bar{t}$ production and single-top production; while the latter process is mediated (at LO) by the electroweak interaction, the former is predominantly induced by the strong force and it is characterized by a larger cross section.

Top quark pair production

The main amplitudes contributing at tree level to the production of a $t\bar{t}$ pair are shown in Figure 1.3(a): these include two different parton-level processes, namely $gg \to t\bar{t}$ and $q\bar{q} \to t\bar{t}$. At the LHC, $t\bar{t}$ production is dominated by the gluon-fusion mechanism, due to the relatively small values of the PDFs of antiquarks in the proton. The exact fraction of the total $t\bar{t}$ cross section due to $gg \to t\bar{t}$ depends on the value of $\sqrt{s}$: at the LHC, this fraction varies from approximately 80% at $\sqrt{s} = 7$ TeV up to roughly 90% at $\sqrt{s} = 14$ TeV [5]. At



(a) (b)

Figure 1.4: Cross section measurements performed by experiments at the Tevatron and LHC for (a) $t\bar{t}$ production and (b) single-top production. Experimental values are shown as a function of the c. o. m. energy of the hadronic collision and compared to theoretical predictions. Both plots are taken from Ref. [5].

the Tevatron, $t\bar{t}$ pairs are predominantly produced via $q\bar{q}$ annihilation. The production of a $t\bar{t}$ pair in QCD can be accompanied by the production of additional partons (quarks or gluons), leading to the presence of additional hadronic jets in the final state. These radiative processes, which are described by higher-order corrections in QCD (see Figure 1.3), become increasingly important at high c. o. m. energies and they can account for a large fraction of the total inclusive $t\bar{t}$ cross section.

Several measurements of the inclusive $t\bar{t}$ cross section have been performed by the Tevatron and LHC experiments at five different c. o. m. energies (see Ref. [5] and references therein). Experimental values are found to be in very good agreement with state-of-the-art theoretical calculations, as shown in Figure 1.4(a). The most accurate theoretical calculations for the inclusive $t\bar{t}$ cross section at the LHC, assuming a top quark mass of 173.2 GeV, yield $\sigma_{t\bar{t}} = 247.7^{+6.3+11.5}_{-8.5-11.5}$ pb at $\sqrt{s} = 8$ TeV and $\sigma_{t\bar{t}} = 816.0^{+19.4+34.4}_{-28.6-34.4}$ pb at $\sqrt{s} = 13$ TeV [5].

Top quark pairs can also be produced in association with other SM particles, e.g. a photon ($t\bar{t}\gamma$), a gauge boson ($t\bar{t}$V with V = W/Z) or a Higgs boson ($t\bar{t}h$). On the other hand, these processes are characterized by very small cross sections, compared to the cross section for the production of a $t\bar{t}$ pair in QCD.

Single-top production

Single-top production at hadron colliders can occur through three different electroweak processes, as shown in Figure 1.5. At the LHC, t-channel production is the process with the largest cross section, followed by the associated production with a W boson (tW) and s-channel production. Experimental measurements of the single-top cross section at the Tevatron and LHC are found to be in good agreement with theoretical predictions, as shown in Figure 1.4(b) [5]. The latest theoretical calculations for the inclusive cross section of single-top production at the LHC predict $\sigma_t = 112.3$ pb at $\sqrt{s} = 8$ TeV and $\sigma_t = 299.0$ pb at $\sqrt{s} = 13$ TeV, assuming a top quark mass of 172.5 GeV [32, 33].

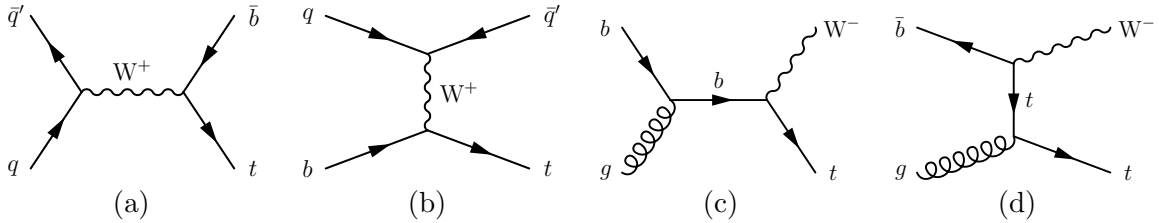


Figure 1.5: Leading-order diagrams for single-top production in a hadron-hadron collision: (a) s-channel, (b) t-channel, (c, d) tW production. Diagrams for single production of an antitop quark are not shown.

1.4.2 Decay channels

Due to the large value of the top quark mass, the phenomenology of top quark decays is unique and very different from that of the other quarks in the SM. The top quark lifetime ($\tau_t \simeq 5 \cdot 10^{-25}$ s) is approximately one order of magnitude smaller than the typical timescale of hadronization ($\tau_{\text{had}} \sim 1/\Lambda_{\text{QCD}} \sim 10^{-24}$ s); as a consequence, the top quark is the only quark that decays before hadronization takes place and its spin information is transferred to its decay products and conserved in their angular correlations.

Since the top quark mass is larger than the mass of the W boson, the dominant decay mode of the top quark is given by the flavor-changing EW process $t \to q_d \mathrm{W}^{\pm}$, where $q_d = d, s, b$ corresponds to one of the three down-type quarks of the SM. As a consequence of the hierarchy of the CKM matrix elements, the top quark decays almost exclusively to a W boson and a bottom quark; the total branching ratio for decays to other down-type quarks ($t \to s\mathrm{W}^{\pm}$ and $t \to d\mathrm{W}^{\pm}$) is expected to be less than 0.2% [5]. The final signature of a top quark decay is thus determined by the decay of the W boson: the latter can decay to a pair of quarks, with $\mathrm{BR}(\mathrm{W} \to q\bar{q}') \simeq 67\%$, or to a charged lepton and a neutrino, with $\mathrm{BR}(\mathrm{W} \to \ell\nu_\ell) \simeq 33\%$.

Decay modes of a $t\bar{t}$ pair and experimental signatures

The decay modes of a $t\bar{t}$ pair can be classified in the following three categories, based on the number of prompt leptons in the final state:

- *all-hadronic* channel, $t\bar{t} \to bq\bar{q}'\ \bar{b}q''\bar{q}'''$ ($\mathrm{BR} \simeq 44\%$),
- *semileptonic* (or ℓ + jets) channel, $t\bar{t} \to b\ell\nu_\ell\ \bar{b}q\bar{q}'$ ($\mathrm{BR} \simeq 15\%$ per lepton flavor), and
- *dilepton* channel, $t\bar{t} \to b\ell^+\nu_\ell\ \bar{b}\ell'^-\nu_{\ell'}$ ($\mathrm{BR} \simeq 11\%$).

The final experimental signature in each case depends not only on the top quark decay products, but also on the momentum of the top quarks produced in the final state. For relatively low values of the top quark transverse momentum ($p_T^t \lesssim 400$ GeV), its decay products are expected to be well distanced in the final state and to be reconstructed in the detector as separate objects. This case is referred to as a top quark *resolved* decay. For higher values of the top quark p_T, the decay products can be highly collimated in the laboratory frame and they can give rise to experimental signatures much different from the resolved case. We refer to this case as to a *boosted* top quark decay. In the case of a boosted semileptonic top quark decay ($t \to b\ell\nu$), the prompt lepton in the final state is not expected to be isolated due to its vicinity to the hadronization products of the bottom quark coming from the

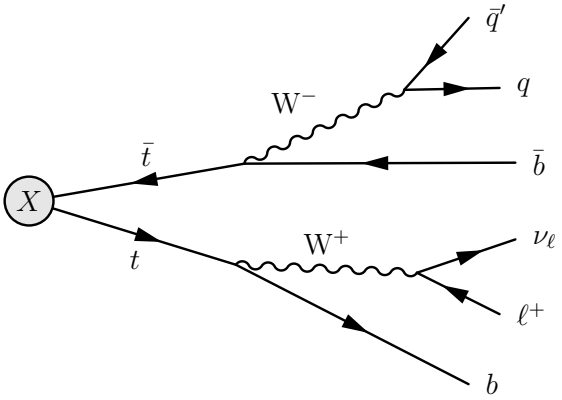


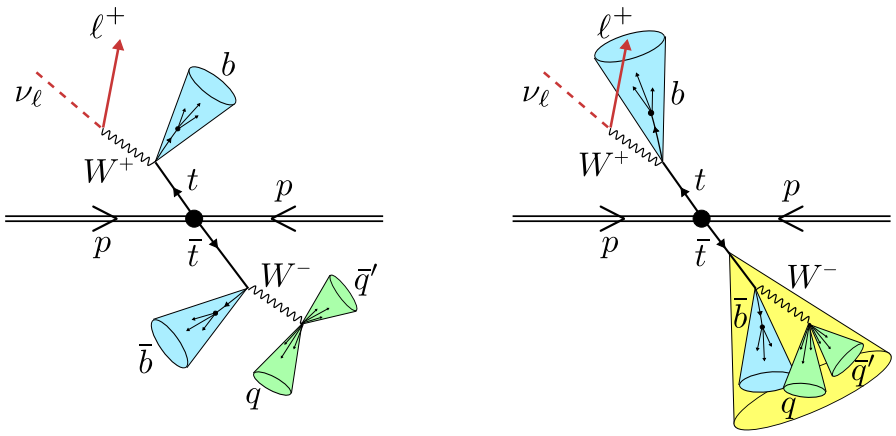
Figure 1.6: Decay chain of a $t\bar{t}$ pair in the ℓ + jets channel.

same top quark decay. Standard requirements like lepton isolation and minimum angular separation between lepton and jets cannot be applied to reconstruct this type of decay. In a boosted top quark hadronic decay, the decay products of the three quarks may overlap significantly and be reconstructed in the detector not as three separated jets, as expected in the resolved case, but as a single large-radius jet characterized by specific substructure properties. The techniques aimed at identifying a merged jet from a boosted top quark decay are known as jet *t-tagging* algorithms [34, 35]. The efficient reconstruction of boosted top quark decays acquires particular relevance in the context of searches for high-mass $t\bar{t}$ resonances ($M_X \gtrsim 1$ TeV): for increasing values of the signal mass hypothesis, the fraction of boosted top quarks produced in the $X \to t\bar{t}$ decay becomes higher and the use of specific methods for the reconstruction of high-p_T top quark decays becomes increasingly valuable.

For both the resolved and boosted regimes, the identification of jets from b-quark decays (jet b-tagging) is in general a very effective method to identify $t\bar{t}$ events in any of the three $t\bar{t}$ decay channels.

In the all-hadronic channel, both W bosons decay hadronically, there are no charged leptons in the final state and no loss of energy due to undetected neutrinos. In the resolved case, the final state is characterized by the presence of six hadronic jets from the $t\bar{t}$ decay. In the case of boosted $t\bar{t}$ production, the decay products of one or more quarks, depending on the top quark transverse momentum, can overlap and be reconstructed in the same jet. For highly boosted $t\bar{t}$ decays, the final state can include as little as just two high-p_T jets, each originating from one hadronically decaying top quark. The biggest challenge for experimental analyses in this channel is the reduction and estimation of the very large background coming from QCD-multijet production. The latter background is difficult to model with MC simulations and a reliable description of it often requires a data-driven method. When reconstructing the kinematics of the $t\bar{t}$ pair in the resolved case, an additional complication is given by the large number of permutations in the assignment of hadronic jets to a given top quark candidate. The problem of jet combinatorics is less important (or even absent) in the boosted regime, as the number of reconstructed jets decreases. On the other hand, in a boosted all-hadronic analysis an additional systematic uncertainty comes from the modeling of the efficiency of the jet t-tagging algorithm.

The ℓ+jets channel, whose decay chain is shown in Figure 1.6, combines a high branching fraction with a relatively clean experimental signature (especially for μ + jets and e + jets), thanks to the presence of one charged lepton in the final state. In the rest of this thesis, we use the expression "ℓ + jets channel" to refer only to the μ + jets and e + jets final states.



(resolved) (boosted)

Figure 1.7: Graphic representation of the $t\bar{t} \to \ell + \text{jets}$ final state in the resolved (left) and boosted (right) regimes.

Semileptonic events with a τ lepton ($\tau + \text{jets}$) are generally studied in dedicated analyses, due to the particular signature of τ decays. The $\ell + \text{jets}$ final state in the resolved regime is characterized by the presence of one isolated lepton, missing transverse energy, and four well-separated hadronic jets. As the two top quarks become increasingly boosted, the final state topology tends to a back-to-back configuration: on one side, the semileptonic top decay produces, in addition to missing transverse energy, one poorly-isolated lepton close in space to a b-jet; on the other side of the event, the boosted hadronic top quark decay leads to the presence of two jets (one from the b-quark decay and one from the $\mathrm{W} \to q\bar{q}'$ decay) or even just one single merged jet comprising all the top quark decay products. Figure 1.7 illustrates the event topologies expected in the resolved and boosted regimes for the $\ell + \text{jets}$ channel. Even though missing energy is expected in the final state due to the presence of one neutrino, the kinematics of the $t\bar{t}$ decay can be fully reconstructed using minimal assumptions. The only ambiguities in the $t\bar{t}$ reconstruction are given by the longitudinal component of the neutrino momentum, which can be inferred for example imposing the W-mass constraint on the $\ell + \nu$ system, and the assignment of jets to the two top quark candidates, which is simplified when the hadronic top quark decay is boosted.

The dilepton channel is characterized by the presence of two opposite-sign charged leptons. This provides a very clean signature at the experimental level with a relatively low rate of non-$t\bar{t}$ backgrounds. This is especially true in the resolved case, while additional complications can occur in the boosted regime due to the difficulty to identify the leptons coming from the top quark decays. The main limitation in this channel comes from its very small branching ratio, which makes it less suited for the discovery of rare new physics phenomena. A second disadvantage in dilepton analyses is that the kinematics of the final state cannot be perfectly reconstructed, due to the presence of two neutrinos. Observables like the invariant mass of the $t\bar{t}$ system can only be estimated in this channel by means of additional assumptions and this, in turn, generally reduces the resolution with which such quantities are measured.

1.5 Previous searches for $t\bar{t}$ resonances at hadron colliders

Searches for $t\bar{t}$ resonances have been carried out in the past with data collected at the Fermilab Tevatron collider and at the CERN LHC. None of these searches has found evidence for the production of a $t\bar{t}$ resonance and they have set exclusion limits on the cross section of resonant $t\bar{t}$ production. In the following, we briefly summarize the results of some of these past analyses by taking as a reference the lower limit set by each experiment on the mass of a leptophobic $Z' \to t\bar{t}$ resonance with a relative width ($\Gamma_{Z'}/M_{Z'}$) of 1.2% in the topcolor model.

The CDF Collaboration at the Tevatron, based on the analysis of the full data set recorded by the experiment in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV, has set upper limits on the production cross section times branching ratio for a narrow resonant $X \to t\bar{t}$ signal and has excluded at 95% CL a leptophobic topcolor Z' boson with a mass lower than 915 GeV [36]. Similar results were found by the D0 experiment in the analysis of data collected at the same center-of-mass energy: the D0 Collaboration has excluded at 95% CL a Z' resonance with relative width of 1.2% (3%) for masses below 835 GeV (940 GeV) [37].

The ATLAS and CMS Collaborations at the LHC have performed searches for resonant $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV, finding no evidence of a BSM signal. The ATLAS experiment has excluded at 95% CL a leptophobic topcolor Z' resonance with a relative width of 1.2% with a mass below 1.74 GeV [38]; for the same benchmark model, the lower mass limit set by the CMS experiment corresponds to 1.49 GeV [39]. Both experiments have set sub-picobarn limits on the production cross section of $X \to t\bar{t}$ resonances with masses between 1 TeV and 3 TeV.

Chapter 2

The CMS experiment at the CERN Large Hadron Collider

The results presented in this thesis are based on the analysis of proton-proton collisions data collected by the Compact Muon Solenoid (CMS) experiment [40] at the CERN Large Hadron Collider (LHC) [41,42]. The LHC is a particle accelerator designed to deliver hadron-hadron collisions up to a center-of-mass (c. o. m.) energy of $\sqrt{s} = 14$ TeV. The CMS experiment is one of the general-purpose experiments designed to analyze the collisions produced at the LHC. Key objectives of the CMS physics program [43, 44] include the study of a very wide range of SM processes, the direct search for the Higgs boson predicted by the SM and, finally, the search for new, yet unobserved, physics phenomena. This chapter is dedicated to a brief description of the LHC machine and the experimental apparatus of the CMS detector.

2.1 The CERN Large Hadron Collider

The Large Hadron Collider (LHC) [41, 42] is a circular hadron-hadron collider operated at the laboratories of the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. The collider has a circumference of 27 km and it is located in the underground tunnel originally constructed to host the Large Electron Positron (LEP) collider [45], which was operated from 1989 until 2000. The main part of the LHC physics program consists in operating the machine as a proton-proton (pp) collider, while part of the machine schedule is periodically dedicated to the delivery of heavy-ion collisions.

The LHC ring includes two adjacent beam pipes, each containing one of two colliding beams, which travel in opposite directions in the collider ring. The two beams are focused and bent in their circular trajectory by a system of more than 1600 superconducting magnets, operating at temperatures between 1.8 K and 4 K (depending on the type of magnet): 1232 dipole magnets operating at a temperature of less than 2 K generate a magnetic field of 8.3 T, which is used to maintain the circular motion of the two beams around the circumference of the LHC; 392 quadrupole magnets are used to keep the beams focused as they travel inside the collider. In the LHC, beams are accelerated by the electromagnetic field generated by radio-frequency cavities (eight per beam) located along the collider ring: each of these cavities also operates in superconducting state, at a temperature of approximately 4.5 K, and can deliver a voltage of 2 MV at a frequency of 400 MHz. Prior to their injection in the LHC, the colliding particles (protons or heavy-ions) are grouped together in bunches and pre-accelerated in a chained system of smaller accelerators, which complete the CERN accelerator complex. A layout of the CERN accelerator complex can be seen in Figure 2.1.

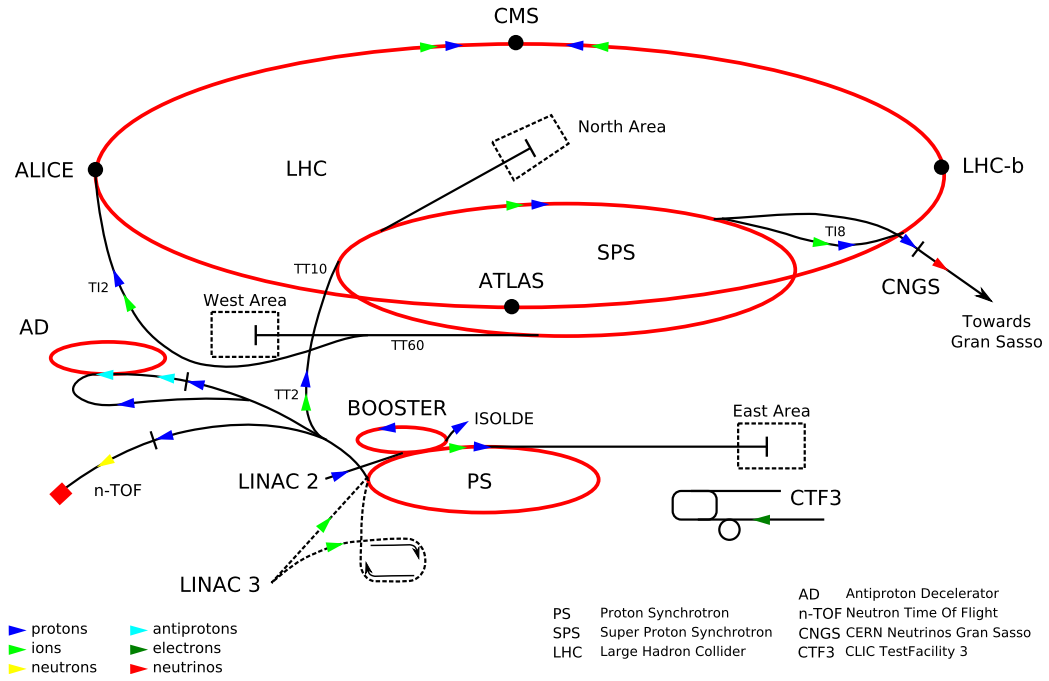


Figure 2.1: Graphic representation of the CERN accelerator complex.

The two collider beam pipes cross at four interaction points, where collisions are delivered to the four multi-purpose experiments operating at the LHC. ALICE (A Large Ion Collider Experiment) [46] is a detector designed for heavy-ion collisions; its main objective is to study the physics of strongly interacting matter at extreme energy densities, including a series of physics phenomena related to the formation of the quark-gluon plasma [47]. The Large Hadron Collider beauty (LHCb) experiment [48] is dedicated to the study of heavy flavor physics: the primary goal of its physics program is to study the properties of CP violation in the quark sector and to search for evidence of new physics in rare decays of charm and beauty hadrons. ATLAS (A Toroidal LHC ApparatuS) [49] and the Compact Muon Solenoid (CMS) [40] are the two largest experiments operating at the LHC. Their extensive physics programs both focus on studying particle physics processes up to energies above the TeV scale, ranging from precision measurements in the context of the SM to the direct search for the SM Higgs and evidence of BSM physics.

In a collider experiment, the production rate of a generic physics process is given by

$$\frac{d\mathrm{N}_p}{dt} = \sigma_p \cdot \mathscr{L} \tag{2.1}$$

where σ_p denotes the total cross section for the process of interest at the center-of-mass energy of the collision and $\mathscr{L}$ is a machine-dependent quantity referred to as the *instantaneous luminosity*, which corresponds to the number of bunch crossings per unit area and per unit time. This quantity can be expressed in terms of a set of parameters related to the beam properties and the geometry of the collisions, according to the following expression

$$\mathscr{L} = \frac{n_b \cdot N_{\mathrm{bunch},1} \cdot N_{\mathrm{bunch},2} \cdot f_{\mathrm{rev}}}{4\pi \cdot \beta^* \cdot \varepsilon_n} \cdot R \tag{2.2}$$

where n_b is the number of proton bunches per beam, $N_{\mathrm{bunch},1/2}$ denotes the number of protons in a single bunch of each beam, f_{rev} is the beam revolution frequency in the LHC ring, β^* is the

insertion region focusing parameter, ε_n is the normalized emittance and R is the interaction region geometric factor. The total integrated luminosity is given by $\mathscr{L}_{\text{int}} = \int \mathscr{L}\, dt$ and it is generally used as a measure of the total amount of data delivered by the machine, or recorded by an experiment, over a certain period of time.

Since a variety of physics processes of interest, e.g. rare SM processes or hypothetical BSM processes, are characterized by very small cross sections, one of the main objectives in the design of a particle collider is to maximize the instantaneous luminosity of the machine in order to reach the sensitivity necessary for such rare processes. On the other hand, the increase of the instantaneous luminosity also implies an increased number of *pp* collisions occurring in the same bunch-crossing; for example, given that the total inelastic cross section for *pp* collisions at $\sqrt{s} = 8$ TeV is approximately 70 mb, an instantaneous luminosity of $4 \cdot 10^{33}$ cm^{-2}s^{-1} and a bunch spacing of 50 ns correspond to an expected number of 14 collisions in the same bunch-crossing. This superposition of multiple collisions, also known as *pileup*, represents one of the main challenges for the LHC experiments in almost all the aspects of their operations, from the design of the online trigger to the performance of the offline event reconstruction in physics analyses. This is due to the higher density of tracks in inner tracking system and contamination in the energy deposits measured in the calorimeters. Pileup interactions are classified as *in-time* pileup, if they originate from the same bunch crossing of the hard-scattering event, or as *out-of-time* pileup, if they come from previous and successive bunch crossings. Signals from out-of-time pileup interactions can affect the read-out of a given detector, if the latter has an integration time longer than the LHC bunch-crossing. The impact of out-of-time pileup can be greatly reduced already at trigger level, combining the timing measurements from different subdetectors. Effects from in-time pileup are more difficult to suppress but they can be mitigated quite effectively as well, especially in the offline analysis of the data.

The LHC was designed to deliver *pp* collisions at a c.o.m. energy of $\sqrt{s} = 14$ TeV with an instantaneous luminosity of $1 \cdot 10^{34}$ cm^{-2}s^{-1}, using a bunch-spacing of 25 ns. In the first years of collisions data-taking, the LHC has been operated below the aforementioned design values, according to a gradual schedule aimed at reducing the time needed for commissioning the machine and to follow a safer strategy concerning the operation of the LHC magnets. The first run of the LHC physics program with *pp* collisions started in 2009. During the course of 2010 and 2011, the LHC delivered *pp* collisions at $\sqrt{s} = 7$ TeV reaching a maximum luminosity of $3.5 \cdot 10^{33}$ cm^{-2}s^{-1}; in this period, the CMS Collaboration collected data for a total integrated luminosity of approximately 5 fb^{-1}. In the 2012-2013 data-taking campaign, which marked the end of the LHC Run-1, the LHC beam energy was increased from 3.5 TeV to 4 TeV per beam and and the machine reached a peak instantaneous luminosity of $7.7 \cdot 10^{33}$ cm^{-2}s^{-1}. In this period the CMS experiment recorded a total integrated luminosity of 19.7 fb^{-1} with all subdetectors fully operational and nominal magnetic field. During the main stages of Run-1, the LHC collided proton beams with a bunch-spacing of 50 ns, corresponding to a maximum of approximately 1400 proton bunches per beam. In the subsequent two years, from the beginning of 2013 up until the first months of 2015, data-taking operations were suspended to allow for upgrades to the LHC aimed at operating the accelerator at an higher c. o. m. energy and at instantaneous luminosities up to $1 \cdot 10^{34}$ cm^{-2}s^{-1}. This period was also used by the LHC experiments for detector maintenance and upgrades. In 2015 the LHC resumed operations at an increased c.o.m. energy of $\sqrt{s} = 13$ TeV, reaching a peak luminosity of $5 \cdot 10^{33}$ cm^{-2}s^{-1} and employing beams with a bunch-spacing of 25 ns; by the end of the first year of the LHC Run-2, the CMS experiment recorded a total integrated luminosity of 2.6 fb^{-1} with a magnetic field of 3.8 T and all subdetectors in full operation.

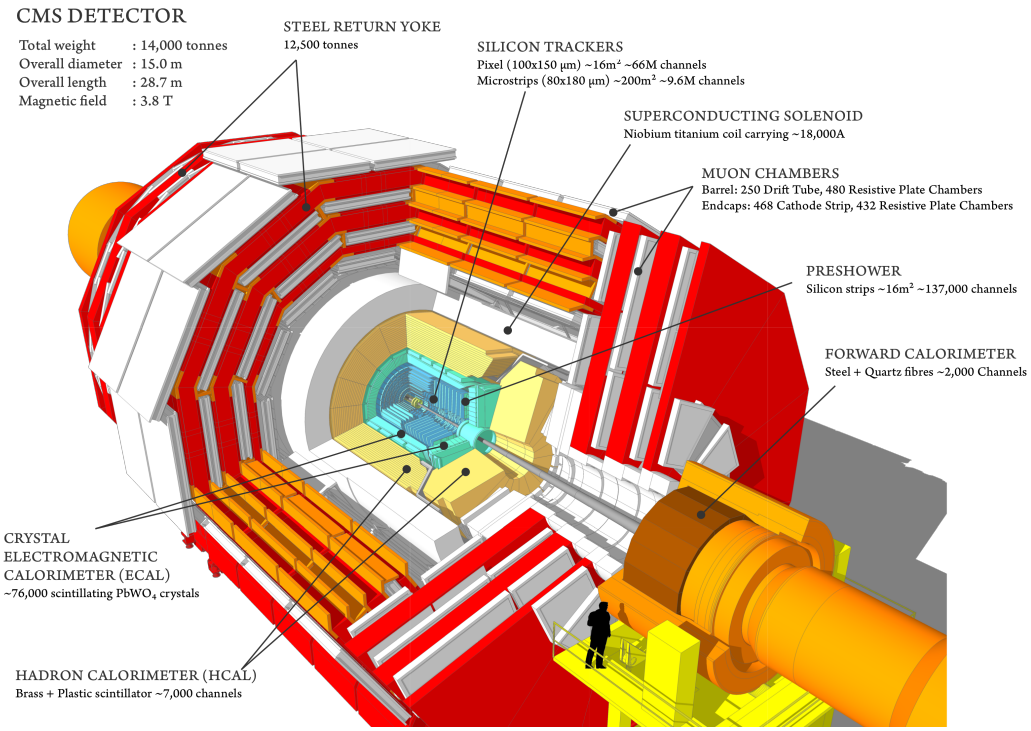


Figure 2.2: Graphic representation of the CMS detector and its subsystems.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) experiment [40,43,44] is one of the two general-purpose particle physics experiments of the LHC complex. The detector is designed to allow for the reconstruction of all types of (detectable) particles produced in the hadron-hadron collisions delivered by the LHC, opening the possibility to perform an extremely wide range of physics measurements, from precision tests of the SM to searches for new physics phenomena.

The general concept driving the detector design is the configuration of the magnetic field needed to bend the trajectory of the charged particles produced in the collision, especially muons, allowing for the accurate measurement of their momentum. The detector hosts, going from the inside out, a tracker, an electromagnetic and a hadronic calorimeter, which are all compact enough to be located inside of the superconducting solenoid. The detectors which comprise the muon system are positioned outside of the magnet coil and they are interleaved with the three layers of the magnet return yoke. Figure 2.2 shows an illustration of the CMS detector and its subsystems.

In the rest of this section, a brief overview of the main subsystems of the CMS detector is given. A detailed description of the CMS detector and its components can be found in Ref. [40] and references therein.

Coordinate system and kinematic quantities

The origin of the coordinate system used by the CMS experiment corresponds to the nominal collision point at the center of the detector. The x-axis points radially towards to the center of the LHC ring, while the y-axis points vertically upwards. The z-axis corresponds to the anticlockwise direction of the beam.

The transverse component of the energy and momentum of a given object with respect to the beam direction, denoted E_T and p_T respectively, are determined from the corresponding x and y components. The imbalance in the total transverse energy of the event, which can be induced by the production of undetected particles, e.g. neutrinos, is referred to as missing transverse energy and it is denoted with the symbol $\not{E}_T$.

The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The polar angle θ is measured from the z-axis. The latter angular position can be expressed in terms of a kinematic quantity called *pseudorapidity* (η), defined as

$$\eta = -\ln\left[\tan(\theta/2)\right] \tag{2.3}$$

In the massless limit ($E \to |\vec{p}|$), pseudorapidity coincides with a quantity known as *rapidity* (y), given by $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$. Rapidity differences, and thus pseudorapidity differences in the massless limit, are invariant under Lorentz boosts along the beam direction. This property is particularly convenient to define the kinematics of particles in a hadron-hadron collision, where the total longitudinal momentum of the parton-parton interaction differs from event to event and cannot be measured. In experimental measurements at colliders, pseudorapidity is preferred to rapidity because the mass of the particle candidate, and thus its total energy, is not known. The angular separation between different physics objects is often measured in the η-ϕ plane, using the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, which is itself invariant under longitudinal boosts.

2.2.1 Magnet system

The main component of the CMS magnet system is a superconducting solenoid based on a niobium-titanium conductor. The solenoid, which is a 13 m-long cylindrical coil with an internal diameter of 6 m, produces a magnetic field of up to 3.8 T along the z-axis, inside the magnet coil. The magnet return yoke in the barrel region is is subdivided along the beam axis into five rings approximately 2.5 m long, with each ring made up of three iron layers. The return yoke in the forward region is comprised of three successive iron disks in each of the two endcaps. In the region outside of the solenoid, the magnetic field has a strength of approximately 2 T and its orientation is inverted, causing the trajectory of muons to be bent in the opposite direction.

2.2.2 Inner tracking system

The innermost subsystem of CMS is the tracking system [50, 51]. The latter is designed to measure the momentum of charged particles stemming from the interaction point by determining the bending of their trajectories in their passage through the detector layers. The individual tracks are also used to reconstruct the event primary vertex and secondary vertices associated to the decay of tau leptons and heavy quarks. Given the high hit rate density expected at the LHC design luminosity and energy (around 1 MHz/mm^2 at 4 cm from the beamline), the tracker design is entirely based on a silicon detector technology, which ensures high granularity, fast response and high radiation tolerance.

The tracking system is composed of a silicon pixel detector in its inner region and a set of detectors based on silicon strips in the outer region. The full coverage of the tracking system extends up to $|\eta| = 2.5$, for a total surface of active silicon of over 200 m^2. The necessity to minimize multiple scattering interactions and energy losses of charged particles in the detector material poses a limitation to the material budget of the tracker. Based

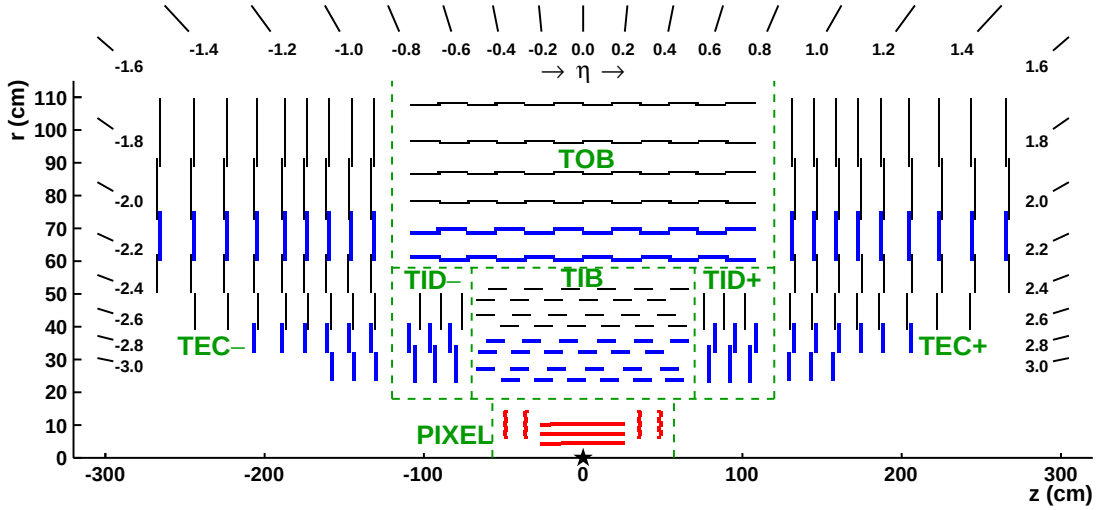


Figure 2.3: Graphic representation of the CMS tracking system in the r-z plane. In this view, the tracker is symmetric with respect to the horizontal line $r = 0$, and only the top half is shown here. Single strip tracker modules are denoted by black segments. The blue segments correspond to two back-to-back strip modules. The pixel modules are shown as red segments [51].

on simulation studies, the detector depth ranges from 0.4 radiation lengths (0.15 nuclear interaction lengths) in the barrel region up to two radiation lengths (0.55 nuclear interaction lengths) in the transition region between barrel and endcap [51]. A schematic view of the CMS tracker is shown in Figure 2.3.

Pixel detector

The pixel detector is comprised of a total of approximately 66 million pixel cells grouped in 1440 modules. Each pixel cell has a size of 100 μm × 150 μm and a thickness of 285 μm. The pixel modules are arranged as follows: in the barrel region, three cylindrical pixel layers are disposed along the beam direction at radial distances of 4.4 cm, 7.3 cm and 10.2 cm from the beampipe; in each of the two endcap regions, two disks of pixel modules are positioned at a distance from the interaction point of approximately 34.5 cm and 46.5 cm in the z-direction. The pixel cells are oriented so that the shorter pitch goes along the azimuthal direction in the barrel and the radial direction in the endcaps. The pixel modules provide a three-dimensional measurement of the hit position, with a typical spatial resolution of 10 μm in the r-ϕ plane and 15 μm along the z-axis, while the third coordinate is given by the sensor plane position. The pixel detector covers a pseudorapidity range $-2.5 < |\eta| < +2.5$, corresponding to the acceptance of the entire tracker. The average efficiency for reconstructing hits is higher than 99%, after excluding a small fraction ($\sim 2\%$) of defective pixel modules [51].

Silicon strips tracker

The outer region of the tracker is based on silicon strip detectors, which can cope with the reduced particle flux coming from the interaction point for an average occupancy (fraction of channels that are hit) up to 2%-3% per strip and LHC bunch crossing. The strip detector is composed of four subsystems, as shown in Figure 2.3. The Tracker Inner Barrel (TIB) and Disks (TID) occupy the region 20 cm $< |r| <$ 55 cm and $|z| <$ 118 cm: the first

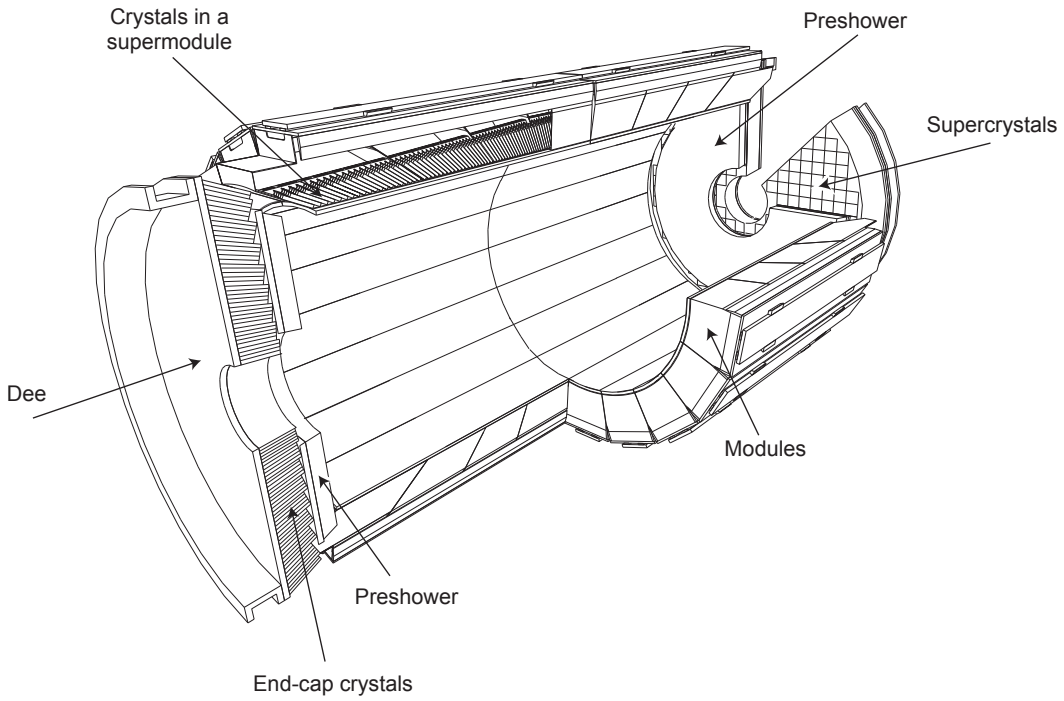


Figure 2.4: Layout of the CMS electromagnetic calorimeter [40].

subdetector is composed by four layers of strip modules in the barrel region, while the second one corresponds to three disks of strips located in the endcaps. The Tracker Outer Barrel (TOB) consists of six barrel layers covering the region with 55 cm $< |r| <$ 116 cm and $|z| <$ 118 cm. Finally, the Tracker EndCaps (TEC) is comprised, on each side, of nine disks positioned along the z-direction (124 cm $< |z| <$ 282 cm). In the barrel, silicon strips are parallel to the beam direction, while in the endcaps they are disposed radially. Each individual strip module provides a two-dimensional measurement of the hit position. The modules in the two innermost layers of both the TIB and the TOB, as well as those in some of the rings of the TID and TEC, are mounted together with a second strip module rotated by an angle of 100 mrad. For each of these back-to-back modules, the measured hits are combined to provide a measurement of the second coordinate along the direction of the strip (z in the barrel, r in the endcaps), thus determining the three-dimensional position of the hit. Strip modules in the TIB and TID provide position measurements with a resolution between 13 μm and 38 μm in $r\phi$. The typical $r\phi$ resolution for the TOB and TEC ranges from 18 μm to 47 μm. The resolution on the position of the second coordinate (r or z), when available, is an order of magnitude worse than in $r\phi$ [51].

2.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [52] is used to measure the energy and position of electrons and photons. The calorimeter technology is based on lead tungstate ($PbWO^4$) scintillating crystals. The relatively high density (8.28 g/cm^3), small radiation length ($X_0 = 0.89$ cm) and small Molière radius (2.3 cm) of the $PbWO^4$ crystals results in a very compact calorimeter with high granularity. These crystals are also characterized by a very fast response (80% of the light is emitted in 25 ns). A representation of the ECAL detector can be seen in Figure 2.4.

The ECAL barrel (EB) is located at a distance $r = 129$ cm from the beam axis and it covers the pseudorapidity range $|\eta| < 1.479$. It consists of a total of 61200 crystals, each with a transverse section of 22×22 mm^2, corresponding to a granularity of 0.0174×0.0174 rad in η-ϕ, and a length of 230 mm (25.8 X_0). Crystals are grouped in arrays of 2×5 crystals, each corresponding to a submodule. Submodules are in turn grouped in modules, each containing 400 or 500 crystals, depending on their position in η. A set of four modules corresponds to a supermodule. Each half of the EB is comprised of 18 supermodules each subtending a 20° angle in ϕ.

The ECAL endcap detectors (EE) is located at a distance $|z| = 315.4$ cm from the interaction point and it covers the range $1.479 < |\eta| < 3$. The endcap crystals have a rear face cross section of 30×30 mm^2, a front face cross section of 28.62×28.62 mm^2 and a length 220 mm (24.7 X_0). These crystals are grouped in 5×5 units, called supercrystals. Each endcap is divided into two semicircular half-disks, each containing a total of 3662 crystals, given by 138 standards supercrystals and 18 partial supercrystals on the inner and outer disk circumferences.

The endcap preshower system is a sampling calorimeter with two layers: in each layer, lead radiators initiate electromagnetic showers from incoming photons/electrons while silicon strip sensors placed after each radiator measure the deposited energy and the transverse profiles of the electromagnetic showers. The preshower detectors are located in front of the ECAL endcaps and cover the pseudorapidity range $1.653 < |\eta| < 2.6$. These detectors are mainly used to identify photon pairs from neutral pion decays. In addition, they allow to improve the position determination of electrons and photons and the identification of electrons against minimum ionizing particles.

The ECAL energy resolution (for energies below 500 GeV) can be parameterized as

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E\ [\text{GeV}]}}\right)^2 + \left(\frac{N}{E\ [\text{GeV}]}\right)^2 + C^2 \tag{2.4}$$

where S is called the stochastic term, N is the noise term and C is the constant term. The stochastic term includes the effects of photostatistics and statistical fluctuations in the shower containment. The noise term quantifies the impact of electronic noise and pileup. The constant term is related to calibration uncertainties and non-uniformities in the calorimeter response. Based on measurements with test-beam data on a barrel module, the typical values of these three parameters are $S = 2.8\%$, $N = 12\%$ and $C = 0.30\%$ [40]; the contribution from the stochastic term, in particular, is considerably small compared to calorimeters used in other collider experiments.

2.2.4 Hadronic calorimeter

The hadronic calorimeter (HCAL) [53] is used to accurately measure the energy of hadrons produced in the LHC collisions. The CMS HCAL is a hermetic sampling calorimeter made of alternating layers of plastic scintillators (active medium) and non-magnetic brass (absorbing material), with the latter having an hadronic interaction length $\lambda_0 = 16.4$ cm. The HCAL system, shown in Figure 2.5, is comprised of four subdetectors which provide an overall $|\eta|$-coverage of $|\eta| < 5$.

The Hadron Barrel (HB) consists of two half-barrel sections covering a total pseudorapidity range of $|\eta| < 1.3$. It is arranged in calorimeter towers, each with a size of 0.087×0.087 in the η-ϕ plane. The HB depth, which is limited by the geometrical dimensions of the ECAL and magnet systems, corresponds to 5.82 interaction lengths in the central region of the

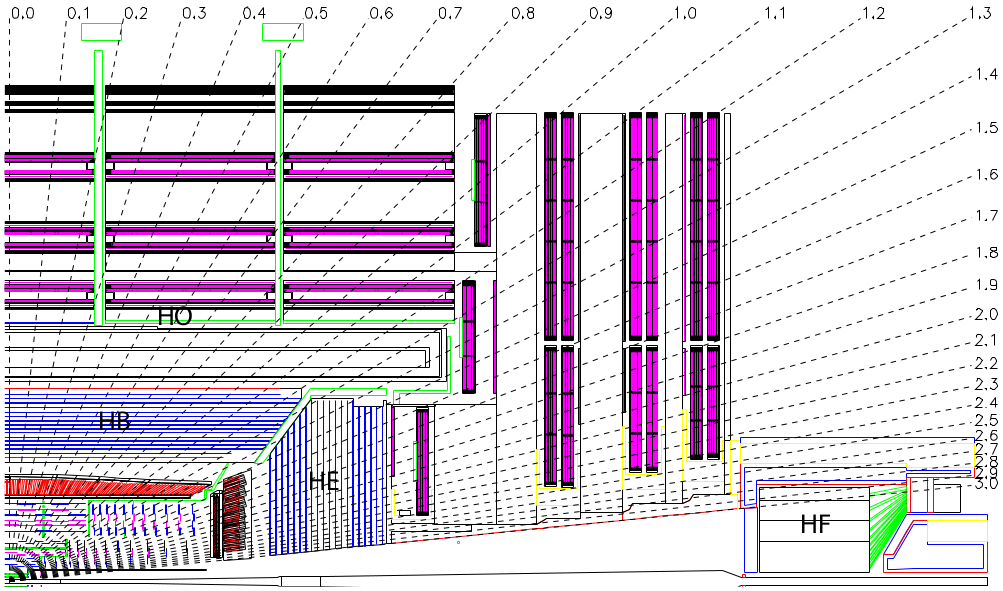


Figure 2.5: Graphic representation of one quadrant of the CMS hadronic calorimeter and its subsystems in the r-z plane [40].

barrel. An additional subsystem, the Hadron Outer (HO) calorimeter, is mounted outside of the magnet coil in the barrel region and it provides additional depth to the HCAL system in the radial direction, up to a minimum of 11.8 interaction lengths (taking into account the solenoid). The segmentation of the HO towers is the same as that of the HB towers.

The Hadron Endcap (HE) calorimeters cover the range $1.3 < |\eta| < 3.0$. Compared to the HB, the segmentation of its calorimeter towers becomes coarser for increasing values of $|\eta|$, with the η-segmentation varying from 0.087 to 0.35 and the ϕ-segmentation ranging from 5° to 10°. This subsystem reaches around $10\,\lambda_0$ in depth (including the ECAL crystals).

The fourth subsystem consists in two Hadron Forward (HF) calorimeters, which cover the very forward region given by $3.0 < |\eta| < 5.0$. These subsystems are located at a distance $|z| = 11.2$ m from the center of the detector. Differently from the other hadron calorimeters, steel is used as absorber in the HF detectors and quartz fibers are used as active medium. This choice is motivated by the higher level of radiation that the detectors are required to withstand in this high-$|\eta|$ region. The segmentation of the HF towers ranges between $0.1-0.3$ in η and $10° - 20°$ in ϕ.

The resolution of the hadronic energy for the combination of the ECAL and HCAL systems can be expressed as

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E\ [\mathrm{GeV}]}}\right)^2 + b^2 \tag{2.5}$$

where a denotes the stochastic term and b is the constant term. Based on test-beam measurements, these parameters have been measured to be $a = 84.7 \pm 1.6\%$ and $b = 7.4 \pm 0.8\%$ in the barrel, with similar values obtained in the endcap region; the corresponding values for the HF calorimeter are $a = 198\%$ and $b = 9\%$ [54].

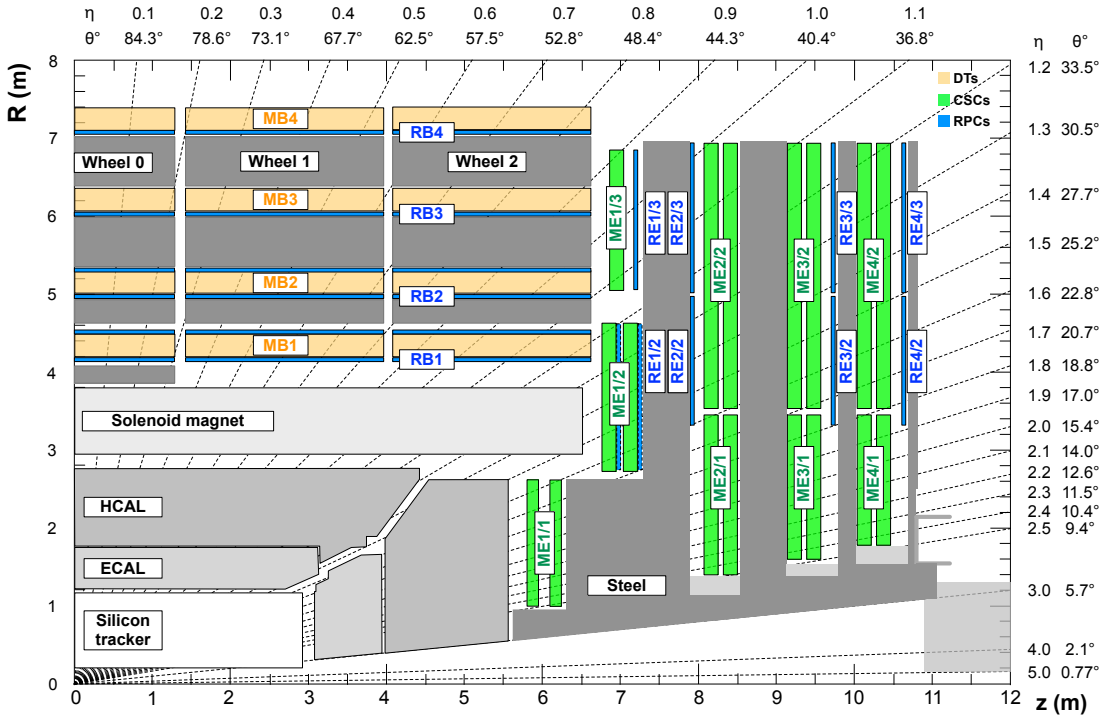


Figure 2.6: Layout of one quadrant of the CMS muon detectors in the r-z plane.

2.2.5 Muon detectors

Muons are distinctive signatures of a multitude of physics processes of interest for the CMS physics program. For this reason, the outer region of the CMS detector hosts dedicated muon spectrometers embedded in the return yoke of the solenoid. Thanks to the strength of the magnetic field, this set of detectors provides a standalone system to efficiently reconstruct muons and determine their transverse momentum, position and electric charge. Measurements in the muon system can also be combined with information from other subdetectors in order to improve the overall performance of muon reconstruction.

The muon system [55], whose layout is shown in Figure 2.6, is located outside of the CMS magnet and it covers a pseudorapidity range equal to $|\eta| < 2.4$. It is comprised of three gaseous detectors based on different technologies: Drift Tube (DT) chambers in the barrel, Cathode Strip Chambers (CSC) in the endcaps and Resistive Plate Chambers (RPC) to complement the first two systems. These subsystems are briefly described in the following.

2.2.5.1 Drift Tube chambers

The barrel region of the muon system is occupied by a collection of Drift Tube chambers. These chambers are allocated in five separate wheels distributed along the z-axis, which cover the pseudorapidity region $|\eta| < 1.2$. In each wheel, four levels (stations) of DT chambers are mounted in the radial direction and interleaved with the layers of the magnet return yoke. Each of the three inner stations contains twelve DT chambers per wheel, while fourteen chambers per wheel are included in the fourth station. The DT chambers in adjacent stations are shifted in the ϕ direction, in order to provide full azimuthal coverage of the barrel muon system and to ensure that a muon coming from the primary vertex crosses at least two chambers. In the three inner stations, each DT chamber is made of three superlayers (SLs);

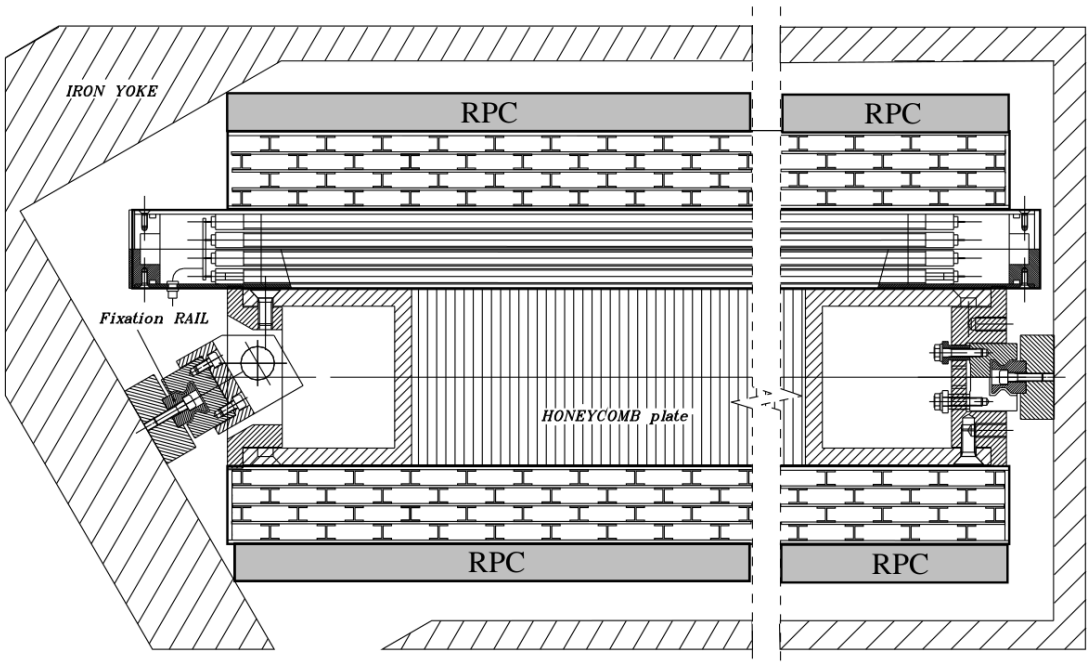


Figure 2.7: Graphic representation of a Drift Tube chamber in the r-ϕ plane, showing two superlayers with wires along the beam direction (ϕ-SLs) and one SL with wires along the ϕ direction (θ-SL) [40].

each SL is comprised of four layers of DT cells disposed in the radial direction and staggered by half a cell. The wires in the first and third SLs are positioned along the z-axis and they are used to reconstruct track segments in the r-ϕ plane; in the intermediate SL, wires placed along the ϕ coordinate are used to measure the track η position. The chambers in the fourth DT station contain only the two ϕ-SLs, so they do not provide a measurement in the z-direction. In each DT chamber, the radial distance between the outer superlayer(s) and the inner one is maximized in order to improve the angular resolution of the final track candidates. The space between SLs is occupied by a honeycomb plate of aluminium which acts as the mechanical support of the DT chamber; in addition, the honeycomb spacer hosts the read-out and trigger electronics that collect the full chamber information. The layout of a DT chamber in the r-ϕ plane is shown in Figure 2.7.

Each Drift Tube cell is based on the design pictured in Figure 2.8. The anode is given by a stainless steel wire located at the center of each cell; aluminium cathodes are located on the cell's sides, while field electrodes are mounted at the top and bottom of the cell. When a muon goes through a DT cell, the ionization of the gas induces of flux of electrons going towards the anode wire: given the linear relationship between the distance from the wire and the electron drift time, the measurement of the latter can be converted into a measurement of the position of the hit in the cell. The cell electrodes are used to shape the electric field inside the cell and improve the linearity between distance and drift time. The spatial resolution for a single-wire measurement is approximately 250 μm in $r\phi$; once multiple hits from the same chamber are combined (up to eight hits in the two ϕ-SLs of a chamber), the resolution for the track segment position per chamber reaches a value between 80 μm and 120 μm in $r\phi$, depending on the wheel and station considered. The chamber resolution for the measurement of the z-position, when available, has a resolution ranging between 140 μm and 390 μm [56]. Track segments measured in each DT superlayer also provide relatively precise

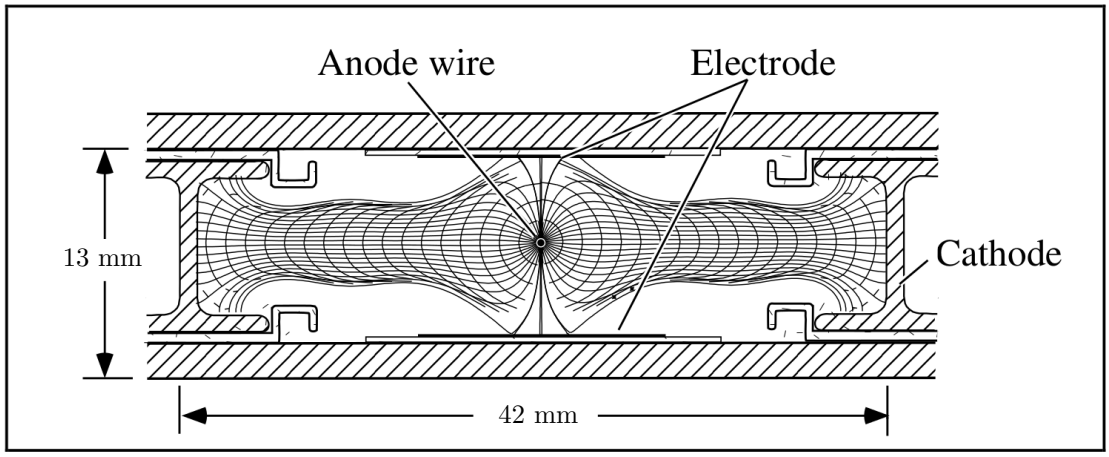


Figure 2.8: Transverse view of a Drift Tube cell; the internal lines correspond to drift trajectories and isochrone lines.

timing measurements, which lead to standalone and efficient bunch crossing identification in the DT system. The time resolution of a local DT segment has been measured to be lower than 2.6 ns [56].

The DT front-end electronics is physically integrated on each DT chamber in an aluminium structure known as *minicrate*. The latter hosts the basic components of the DT read-out and trigger electronics. Each minicrate contains between five and seven Read Out Boards (ROBs), which handle the transmission of DT data to the next stages of the data acquisition chain. For a description of the DT trigger electronics we refer to Section 2.2.6.1.

2.2.5.2 Cathode strip chambers

Cathode Strip Chambers are used for the muon endcap detectors ($0.9 < |\eta| < 2.4$). Compared to the barrel region, this region of the detector is characterized by an higher muon rate, a larger amount of neutron background and the higher intensity and non-uniformity of the magnetic field. The choice of using CSCs is thus motivated by their fast time response, radiation tolerance and fine segmentation.

In each endcap, chambers are arranged in four disks perpendicular to the beam direction; the innermost disk contains three concentric rings, while the others contain only two. Each rings is composed by multiple trapezoidal chambers, each covering an angle between 10° and 20° in ϕ. A Cathode Strip Chamber, shown in Figure 2.9, is a multiwire proportional chamber comprised of six wire chambers (anode) interleaved among seven panels of copper strips (cathode). The anode wires are disposed along the azimuthal direction and identify the radial coordinate of a track hit. The cathode strips are oriented radially, almost perpendicularly to the wires and they provide a measurement of a hit's position along the wire direction (ϕ); the latter measurement is obtained by interpolating the charge distribution induced on the strips by the passage of a charged track (see Figure 2.9). Based on their fast response, the wires are also used to measure the timing of the track hits. The position measurements in each CSC have a typical resolution of 200 μm in r and between 40 μm and 150 μm in $r\phi$ [56]. The time resolution of a CSC segment for muons with $p_T > 20$ GeV in collision events is measured to be approximately 3 ns [56].

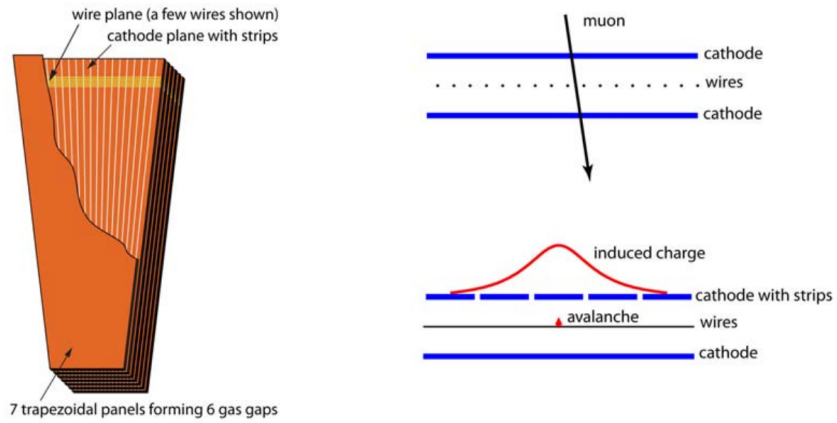


Figure 2.9: Left: layout of a Cathode Strip Chamber. Right: Schematic of the principle of CSC operation, based on charge interpolation on the cathode strips [40].

2.2.5.3 Resistive plate chambers

The third subsystem of the muon detector is comprised of Resistive Plate Chambers. These are gaseous parallel-plate detectors that combine adequate spatial resolution with very precise time measurements. RPCs are mounted on top of the DT chambers and CSCs in the pseudorapidity region $|\eta| < 1.6$ and they are used to complement the information recorded by these other two subsystems. Compared to the DT and CSC detectors, the RPC provide a much improved timing resolution, but an inferior resolution in momentum and spatial position (between 0.8 cm and 1.2 cm [56]).

The basic module of a RPC consists of a double-gap chamber. In each gap, the gas mixture is contained within two parallel plates operated as electrodes in avalanche mode. Read-out strips are mounted between the two chamber gaps and they are used to measure time and position of hits induced by the passage of a charged particle through the chamber. The chamber strips are disposed along the beam direction in the barrel RPCs and in the radial direction in the endcap RPCs.

The time resolution of a RPC, which is of the order of 1 ns, is much smaller than the interval between two consecutive LHC bunches (25 ns); based on this, the main functionality of this subsystem consists in measuring the bunch-crossing associated to a given muon hit and in providing this information to the trigger system. In addition to timing, the trigger of the RPC detector also combines hits measured in separate chambers to determine the trajectory, and thus the transverse momentum, of a track candidate.

2.2.6 Trigger and data acquisition

The design bunch spacing interval of 25 ns in the LHC proton beams corresponds to a bunch-crossing rate of 40 MHz. In addition, depending on the instantaneous luminosity, multiple pp collisions can occur in the same bunch-crossing (approximately twenty at the LHC design luminosity). Because of the inherent limitations in the performance of the detector's read-out and the capability of the storage system, the overall event rate of the detector output must be reduced to a few hundreds of events per second. The detector component dedicated to this task is the trigger system [57, 58], which selects in real time the subset of events to be recorded on disk, based on their physics properties.

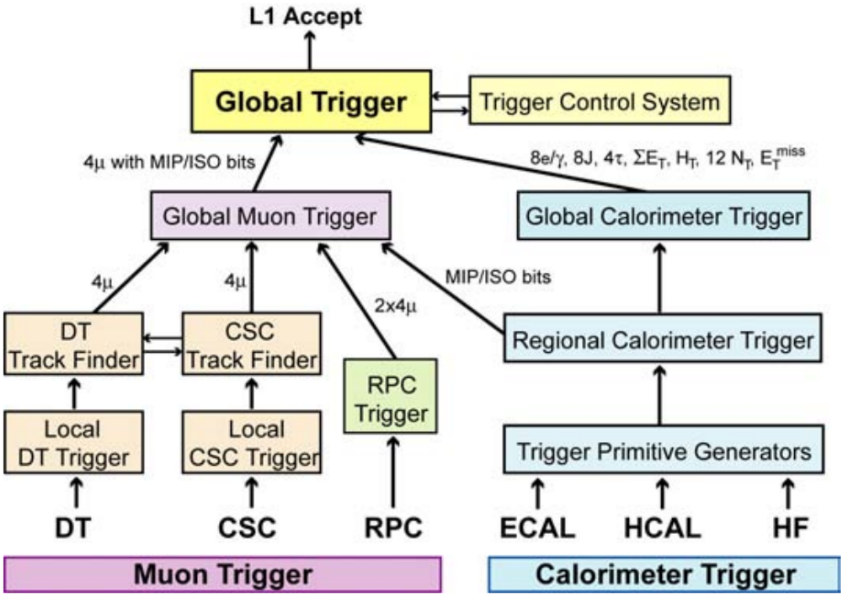


Figure 2.10: Schematic of the CMS L1-Trigger architecture [40].

Due to the large rate reduction factor the trigger is required to achieve and the necessity to maintain the highest possible efficiency to identify the events with interesting physics signatures, the trigger design is based on a two-level architecture. The first layer of the trigger system is the Level-1 (L1) trigger, which uses minimal information from a subset of the CMS subdetectors to perform the first level of the online event selection. The second trigger level, called High Level Trigger (HLT), processes the output of the L1 trigger and it utilizes more sophisticated reconstruction algorithms to ultimately determine if the event is to be recorded.

2.2.6.1 Level-1 trigger

The CMS L1-Trigger is comprised of custom-built hardware processors and it is designed to reduce the initial event rate to less than 100 kHz.

Due to the impossibility to read out the tracker information in this first stage, the L1-Trigger uses only measurements with lower resolution from the muon detectors and the calorimeters. The entire event read-out is temporarily stored in dedicated buffers for a maximum time period of 4 μs, before being discarded or transmitted to the HLT based on the L1 decision. Both the muon trigger and the calorimeter trigger are organized in local, regional and global triggers. Local triggers are used to identify object seeds, called trigger primitives, in each component of a given subdetector. Regional triggers combine the information of the local triggers to construct higher-level L1 objects, e.g. muons, electrons and jets, and estimate some of their basic properties, for example their transverse momentum. The global muon trigger and the global calorimeter trigger rank their respective regional trigger objects based on kinematics and reconstruction quality. The best trigger candidates for each event are sent to the Global Trigger (GT), which finally selects or rejects the event based on a set of programmable trigger requirements. A simplified diagram of the L1-Trigger architecture is shown in Figure 2.10.

Calorimeter trigger

The primitives of the local calorimeter trigger are obtained from the signals coming from the calorimeter towers. These include the measurements of the transverse energy determined in ECAL crystals and HCAL towers and the bunch crossing assigned to the calorimeter signals. Each trigger tower has a (η, ϕ)-coverage of 0.087×0.087 in the region $|\eta| < 1.74$ and a larger size outside of that region. The front-end electronics used for the reconstruction of the trigger primitives is integrated with the on-detector calorimeter read-out.

The calorimeter primitives are transmitted to the Regional Calorimeter Trigger (RCT). which determines for each calorimeter region electron/photon candidates, energy sums and information relevant for muon trigger candidates related to muon isolation (ISO bits) and compatibility with minimum ionizing particles (MIP bits). A calorimeter region corresponds to a set of 4×4 trigger towers, except for the HF subsystem where a single trigger tower corresponds to a region. The electron/photon trigger algorithm determines four isolated and four non-isolated e/γ candidates for each calorimeter region. The RCT also determines τ-veto bits for the identification of narrow hadronic jets coming from the decay of a τ lepton.

The data from the RCTs are forwarded to the Global Calorimeter Trigger (GCT). The GCT determines the best-ranked isolated and non-isolated e/γ trigger candidates and it combines the energy sums from separate calorimeter regions in order to estimate the total missing transverse energy and the total transverse energy in the event. At this stage, a clustering algorithm is applied to regional calorimeter signals to determine jet candidates and jet multiplicities at trigger level. The GCT ultimately selects up to four jets and four τ-jets in the HCAL, in addition to four jets from the HF calorimeter.

Muon trigger

The L1 Muon Trigger processes information coming from all three subdetectors of the muon system (DT, CSC and RPC), in the form of track segments, hit patterns and timing, and it combines it in order to determine the four best muon candidates to be sent to the GT.

The electronics of the DT local trigger is hosted in the minicrates mounted on each DT chamber. The Bunch and Track Identifier (BTI) unit generates a trigger signal based on the wire response and it determines the position and angular direction of track segments in each chamber SL, based on the association of hits in the four planes of the SL. The Track Correlator (TRACO) unit performs a spatial matching of ϕ-segments reconstructed at the same bunch crossing in the two ϕ-SLs of the chamber. If a match is found, the TRACO produces a new segment with improved angular resolution and it assigns a quality score to it. Every muon chamber contains up to twenty-five TRACOs: each of them selects a maximum of two track segments per bunch crossing and sends them to the Trigger Server (TS) units. Each TS has two subsystems. The first one (TSϕ) analyzes the segments measured in the transverse plane by the TRACO, while the second subsystem (TSθ) processes directly the BTI output from the θ-SLs. The trigger information collected from each DT sector is managed by the Sector Collector unit. These units provide the inputs to the DT regional trigger, the Drift Tube Track Finder (DTTF). In the local trigger of the CSC detectors, muon track segments, also called Local Charged Tracks (LCT), are first determined separately by CSC anode wires and cathode strips, for each of the six detector layers. The cathode electronics is designed to measure the ϕ coordinate, while the anode electronics is used to identify the bunch crossing. Cathode and anode signals are then correlated in time and in the number of hit layers in order to determine the final track segments. Up to three LCTs per station are sent to the CSC regional trigger, the CSC Track Finder (CSCTF). The RPC

detectors are primarily used for the purpose of triggering. Thanks to a timing resolution of the order of 1 ns, they provide a precise identification of the event bunch crossing. The RPCs also provide an estimate of the muon candidate transverse momentum by combining ϕ measurements of hits in successive layers and determining the trajectory of the particle under the magnetic field. The RPC muon candidates are sorted separately in the barrel and forward regions and the four best candidates in each region are sent directly to the Global Muon Trigger (GMT).

The regional muon trigger is comprised of the Drift Tube Track Finder in the barrel region and the CSC Track Finder in the endcaps. Each of these subsystems delivers up to four muon candidates to the GMT, ranked based on the muon candidate's transverse momentum and quality. The algorithm used by the DTTF for track reconstruction relies on the extrapolation of a track segment in one muon station to a possible matched segment in another station, under the assumption that both segments belong to a charged particle originated at the interaction vertex. Compatible track segments from a maximum of four muon stations are combined to form a complete track candidate and calculate its transverse momentum, position and quality. A processor called Phi Track Finder (PHTF) performs the reconstruction of muon tracks in the r-ϕ plane for each DT sector. Each PHTF receives a maximum of two track segments per DT chamber and tries to match them. The matching procedure relies on an extrapolation principle based on the correlation between the azimuthal difference of the segments ($\Delta\phi$) and the segment bending angle (ϕ_B). Compatible segments are combined to form tracks and assign to them a quality code. The two track candidates with highest quality are also assigned physical parameters such as transverse momentum, ϕ-position and electric charge. The PHTF processors are allowed to exchange data, in order to reconstruct muon tracks with segments in separate sectors. This in turn demands the application of a cancellation scheme to remove duplicated tracks. The assignment of η-values to the track candidate is performed by the Eta Track Finder (ETTF) processors by matching the hit pattern in the θ-SLs to the trajectory of the PHTF track candidate. The combined output of PHTF and ETTF for each sector consists of at most two track candidates. The Wedge Sorter (WS) unit then determines the best two candidates for every wedge of the barrel system, where a wedge corresponds to the five sectors (from different wheels) at the same ϕ-position. The twenty-four muon candidates selected by the WSs are finally delivered to the Barrel Sorter (BS), which selects the best four candidates in the barrel muon system and sends them to the GMT. Figure 2.11 shows a schematic of the DTTF trigger system. Similarly to the DTTF, the CSCTF applies an extrapolation procedure to the three-dimensional spatial measurements provided by the CSC chambers. Complete tracks are determined from the extrapolation results and redundant tracks are removed. Up to three muon candidates are determined for each of the twelve ϕ-sectors of the CSC system. Finally, the CSC four best muon candidates, selected based on p_T and quality score, are delivered to the GMT.

The Global Muon Trigger (GMT), the last element of muon trigger logic, is designed to improve the trigger efficiency and minimize the trigger rate by exploiting the complementarity and redundancy of the three muon subdetectors. After every LHC bunch crossing, the GMT receives up to four muon candidates each from the DTs, the CSCs, the barrel RPCs and the endcap RPCs. These are combined with information coming from the Global Calorimeter Trigger related to the isolation deposits and compatibility with minimally ionizing particles. If possible, muon candidates from the RPCs are merged with compatible candidates from the DT and CSC regional triggers. The GMT finally selects the four best muon candidates based on the their transverse momentum and quality.

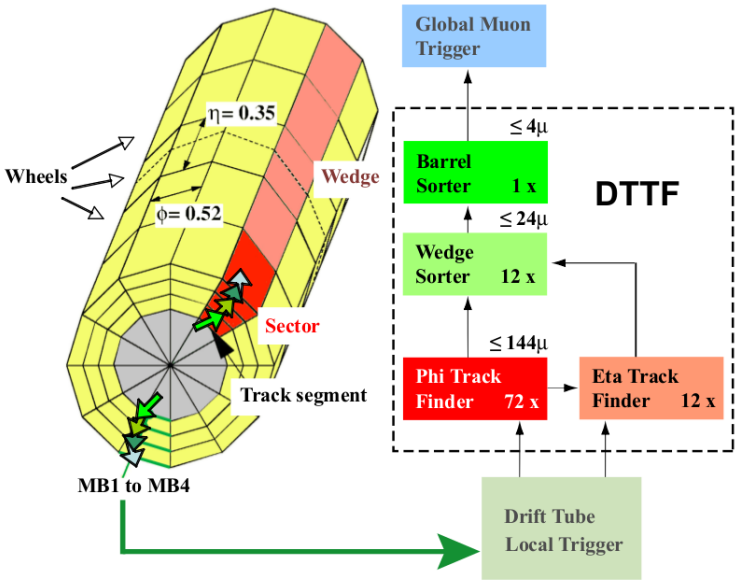


Figure 2.11: Layout of the Drift Tube Track Finder of the muon trigger [59].

The CMS group at the Universidad Autónoma de Madrid (UAM) is responsible for the design, production, commissioning and operation of the L1 DTTF muon trigger.

Global trigger

The Global Trigger is the final stage of the L1 trigger. It receives the trigger objects reconstructed by the GCT and GMT and determines whether the event should be rejected or passed to the HLT. The L1 trigger objects received by the GT include isolated and non-isolated e/γ candidates, muon candidates, central and forward jets, τ-jets and global event quantities such as $\not{E}_T$ and H_T. Multiple algorithms are implemented in the trigger logic in order to select events based on single-object and multi-object requirements on the L1 candidates. Once the outcome of the L1 decision is determined, this information is transmitted back to the subsystems to manage the read-out of the detector data.

2.2.6.2 High level trigger

The High Level Trigger (HLT) [60] further reduces the event rate delivered by the L1-Trigger ($\sim$ 100 kHz) down to an average rate of 400 Hz for offline storage.

This trigger system consists in a farm of commercial processors and it uses reconstruction algorithms which are very similar to the ones used for the offline event reconstruction. The HLT has access to the entire read-out information of the detector, but its logic is optimized in order to minimize the processing time for each event and use the full detector read-out only when necessary. For this reason, the HLT selection proceeds in two successive stages. In the first stage (Level-2), data from the muon and calorimeter detectors is used to reconstruct particle candidates with increased precision compared to the original L1 candidates and a first event selection is used to reduce the input rate from the L1 trigger. For the events accepted to the second stage (Level-3), the full information from the inner tracking detectors is accessed in order to perform the reconstruction of tracks and vertices and apply more refined selection requirements. The typical HLT processing time for an event is around 100 ms.

Data processing at the HLT level is structured around the concept of *HLT path*. The latter corresponds to a sequence of object-reconstruction and event-selection algorithms designed to identify events with specific physics properties. Events selected by at least one HLT path are retained for permanent storage. In particular, triggered events are grouped together in non-exclusive *data-streams*, each corresponding to the combined output of HLT paths targeting similar event signatures. Examples of such data-streams are those based on single-muon and single-electron trigger paths.

2.2.6.3 L1T and HLT performance

The performance of the L1-Trigger and High-Level Trigger systems has been studied in detail for individual physics objects and a wide range of final state signatures [61].

- L1 electron/photon trigger performance: the E_T-resolution of e/γ candidates reconstructed at L1, measured with respect to offline reconstructed electrons, ranges between 10% and 15% depending on the electron η. The L1 electron trigger efficiency for a E_T-threshold of 15 GeV on the L1 EG candidate reaches 99% for electrons with an offline reconstructed E_T of 28.0 ± 0.1 GeV in the barrel and 30.8 ± 0.5 GeV in the endcaps. The reconstruction of electron candidates at HLT is performed with a Kalman filter technique which is complemented, when possible, by the GSF algorithm, which better parameterizes the highly non-gaussian distribution of electron energy losses. The electron HLT efficiency, measured in a sample of $Z \to e^+e^-$ events, reaches a plateau value higher than 99% for electron candidates passing a standard offline selection.

- L1 muon candidates reconstructed in the barrel region have an overall p_T-resolution of $\delta p_T = 14\%$, a ϕ-resolution of $\delta\phi = 0.02$ rad and a η-resolution of $\delta\eta = 0.02$ with respect to offline reconstructed muons. The overall GMT efficiency, measured in data for muons coming from J/ψ and Z boson decays, reaches typical plateau values around 95%. The combined L1+HLT efficiency for single-muon triggers, with and without isolation requirements, reaches typical plateau values ranging between 85% in the endcaps and 95% in the barrel region.

- The performance of L1 single-jet triggers is measured in data using a sample defined by a single-muon trigger, which is independent with respect to jet trigger paths. The turn-on curves determined as a function of the offline jet p_T shows that a plateau L1 efficiency above 99% for a L1 p_T-threshold of 16 GeV (36 GeV) is reached for offline jet p_T values above 20 GeV (60 GeV). At the HLT level, a single-jet HLT path with a p_T threshold of 128 GeV reaches an efficiency close to 100% for offline-p_T values above 170 GeV. These turn-on curves are only mildly dependent on the number of pileup interaction, for the typical pileup multiplicities produced at the LHC up to date. Dedicated algorithms are employed both at L1 and HLT to identify jets compatible with the hadronic decay of τ leptons, yielding a typical overall trigger efficiency as high as 90%. In the HLT stage, tracking information is also used to identify jets originating from b quark decays; for a mistag rate of 1%, the online b-tagging selection at HLT leads to a signal efficiency close to 60%.

In addition to considering the performance of single-object and multi-objects triggers, the performance of the CMS trigger has also been studied for the full trigger selections used in several physics analyses. These triggers often correspond to more complicated algorithms in which requirements are applied on different types of objects and physics quantities. These

measurements confirm the excellent performance of the trigger system in efficiently collecting data for a variety of relevant physics analyses, ranging from B-physics to searches for BSM phenomena [61].

2.2.7 Luminosity monitors

The CMS experiment utilizes some of its subdetectors to measure directly the instantaneous luminosity recorded by the experiment, for the purpose of both online monitoring and offline analysis. In general, luminosity measurements are based on measuring the cross section of very inclusive physics processes, e.g. the inclusive pp cross section or the total cross section for the production of a W or Z boson. The advantage of measurements based on the total pp cross section is that they only rely on simple hardware without requiring all the functionalities of the CMS detector.

The two principal CMS subsystems used as luminosity monitors are the Hadron Forward calorimeter and the silicon pixel detector. Two methods can be used to extract a real-time relative instantaneous luminosity with the HF calorimeter. The first is based on measuring the average fraction of empty towers to infer the mean number of interactions per bunch crossing. The second method exploits the linear relationship between the average transverse energy per tower and the instantaneous luminosity. Luminosity measurements performed with the pixel detector are based on the pixel-cluster counting method [62], which relies on the fact that the number of hit pixel clusters per crossing is a linear function of the number of interactions per crossing.

The absolute calibration of the luminosity measurement is obtained by means of the so-called "van der Meer scan" technique [63]. In the latter, the two LHC beams are scanned across each other in the transverse plane to determine the beam profiles and extract the value of the instantaneous luminosity from machine parameters.

2.2.8 Software and computing infrastructure

The CMS Software (CMSSW) [43] provides the basic software architecture for a very wide range of functionalities which are instrumental to the experiment's goals. These include the algorithms used online for the HLT and the data-quality monitoring, the software tools for the alignment and calibration of the detector and the modules for the final offline reconstruction of the events for physics analyses. The CMSSW framework is also integrated with multiple Monte Carlo (MC) generators in order to produce simulated samples of different physics processes, up to the full simulation of the detector response.

The CMS experiment uses a distributed computing model [64], which utilizes the resources of the Worldwide LHC Computing Grid (WLCG) [65]. The WLCG is structured in three levels (tiers). The main computing center (Tier-0) is a CPU farm located at CERN, in charge of the prompt processing of the raw data coming from the detector and the creation of the datasets. All the other computing centers are distributed in different locations around the world. The Tier-1 computing centers are employed to distribute the datasets and perform event reconstruction, calibration and other computing-intensive tasks. They also provide more compact versions of the original datasets, which can be allocated to smaller sites, i.e. the Tier-2 and Tier-3 centers. The latter provide additional data storage and CPU power for data analysis and event simulations.

Chapter 3

Reconstruction of physics objects

This chapter describes the basic methods used for the offline reconstruction of physics objects in the CMS experiment. These methods represent the building blocks of the physics analyses discussed in later chapters. These analyses rely on the full-event reconstruction provided by the CMS Particle Flow algorithm, which is introduced at the beginning of this chapter. This is followed by a description of the reconstruction methods employed for individual physics objects, such as muons, electrons, jets and missing energy[1]. The chapter also includes an overview of the methods used to identify hadronic jets associated to the decay of a bottom quark (jet b-tagging) and a top quark (jet t-tagging). These reconstruction techniques play a crucial role in physics analyses targeting final states characterized by top quark production.

3.1 Particle Flow algorithm and Global Event description

The Particle Flow (PF) algorithm [66] is a method designed to reconstruct all detectable particles produced in a collision event by combining all the available information from the subdetectors of CMS. This approach leads to a global reconstruction of the event, which is described as a collection of mutually exclusive PF candidates. Each of these candidates is identified as either a muon, an electron, a photon, a charged hadron or a neutral hadron. One of the advantages of the PF method is that the precise reconstruction of all the individual final state particles, including the ones with relatively low p_T, leads to an improved performance in the reconstruction of higher-level objects, e.g. jets, $\not\!\!E_T$ and hadronic τ decays.

The feasibility of this global-event reconstruction relies on the capabilities of the CMS detector, in particular the strong bending power provided by the CMS magnet, the high granularity of the inner tracking system and ECAL, the hermeticity of the HCAL and the high performance of the muon spectrometers. The PF algorithm was first commissioned in CMS with early collisions data taken in 2009 [67]. Further measurements confirmed that the PF-based event description leads to an improved performance in the reconstruction of all physics objects, in particular jets and $\not\!\!E_T$ [68]. Dedicated studies on the performance of the PF algorithm for muon and electron reconstruction have also been carried out [69].

The PF algorithm consists in three successive steps. The first stage is concerned with the reconstruction of the basic PF elements in the different subdetectors, namely tracks in the inner tracker, calorimeter clusters and tracks in the muon system. Secondly, these basic components are associated to one another via the so-called *link algorithm*. Lastly, individual particles are reconstructed and identified.

[1] The methods for the reconstruction of photons and τ leptons are not detailed in this chapter, since they are not employed in the physics analyses discussed in this thesis.

Tracks and calorimeter clusters

Tracks in the inner tracker are reconstructed using the iterative tracking method described in Section 3.2. Calorimeter clusters are identified separately for each subsystem of the ECAL and HCAL detectors. Muons tracks in the muon spectrometers are determined using the standard methods for muon reconstruction (Section 3.3).

Link algorithm

Each type of particle gives rise to a specific signature in the detector, which can comprise one or more of the basic objects described above, i.e. tracks and clusters. For example, a charged hadron is expected to produce a track in the inner tracker and energy deposits in the ECAL and HCAL detectors. For this reason, the PF reconstruction uses a *link algorithm* to create a correspondence between basic objects measured in different subdetectors. The association utilized in the linking stage is purely geometrical. Tracker tracks are extrapolated to the calorimeters and matched to a cluster if the extrapolation falls within the cluster boundaries. If a track matches more than one cluster or vice versa, all combinations are retained. In the case of electrons, candidate bremsstrahlung clusters in the ECAL are associated to a track if their position is compatible with the tangent of the track trajectory extrapolated from a tracker layer to the ECAL surface. A cluster in the ECAL is linked to a cluster in the coarser HCAL if the former lies within the cluster envelope of the latter. Finally, tracks in the muon spectrometer are linked to compatible tracks in the inner tracking system.

Particle identification

Linked PF elements are used to reconstruct and identify the PF candidates in the event. This identification proceeds in sequential steps. The PF elements associated to the particles identified in each stage are removed from further consideration, avoiding any double-counting of the PF elements reconstructed in the event.

First, muon candidates passing either a detector-level isolation or tight identification requirements are identified as muons. Electrons candidates are identified based on the properties of the corresponding track and ECAL clusters, as described in Section 3.4. The remaining tracker tracks associated to calorimeter clusters are identified as charged hadrons. If the momenta estimated for the track and the cluster are found to be compatible, the two measurements are combined to provide the best estimate of the charged hadron energy. In the case where the track momentum is found to be significantly higher than the one expected for the cluster, the particle is marked as a muon if it satisfies a set of loose muon identification cuts. If the cluster energy significantly exceeds that of the associated track, the residual energy may be assigned to photons (ECAL clusters) and neutral hadrons (ECAL and HCAL clusters). Finally, additional photons and neutral hadrons are determined from calorimeter clusters not associated to any tracks.

3.2 Reconstruction of tracks and primary vertices

The reconstruction of tracks in the inner tracking system is one the most important components for the reconstruction of physics objects in CMS. Track reconstruction serves as one of the basic elements of the PF algorithm and it is used to determine the position of the primary vertices in the event and separate the hard-scattering process from pileup interactions.

Track reconstruction

The tracking algorithm used in the CMS experiment is known as the Combinatorial Track Finder (CTF) [51]. The latter represents an extension of the Kalman Filter [70]. The track reconstruction procedure consists in multiple iterations of the CTF algorithm, in a process known as *iterative tracking*. In the first iterations, tight criteria are used to identify the cleanest tracks near the beamspot position. Looser requirements are applied in later iterations in order to reconstruct more complex trajectories associated, for example, to low-p_T particles and displaced tracks. The tracker hits associated to the tracks determined in each iteration are not considered in the subsequent stages, reducing the complexity of the hit combinatorics and thus limiting the track mis-identification rate. Each iteration can be divided in four separate steps.

- Seed generation: an initial estimate (seed) of the track trajectory is determined based on a limited number of hits in the silicon detectors.
- Track finding (or pattern recognition): the seed trajectory is extrapolated to the outer layers of the tracker and new hits compatible with the original track are found.
- Track fitting: the final collection of hits associated to the track are fitted in order to determine the best estimate of the track parameters; after this, spurious hits (outliers) with limited compatibility with the track trajectory may be removed from the track.
- Track selection: track candidates are retained only if they satisfy a set of track-quality requirements.

The performance of track reconstruction in CMS has been studied extensively using collisions data and simulated events [51]. The tracking efficiency (fraction of particles with a reconstructed track associated to it) in data for isolated muons with 1 GeV $< p_T <$ 100 GeV is higher than 99% over the entire η-range covered by the tracker. For simulated $t\bar{t}$ events under typical pileup conditions, the average track-reconstruction efficiency for promptly-produced charged particles with $p_T > 0.9$ GeV is 94% for $|\eta| < 0.9$ and 85% for $0.9 < |\eta| < 2.5$. The corresponding fake rate corresponds to approximately 3% in the barrel region ($|\eta| < 0.8$), and it reaches a maximum value of 10% in the barrel-endcap transition regions.

The transverse momentum resolution for inner tracks associated to isolated muons with $p_T \simeq 100$ GeV and $|\eta| < 1.4$ is approximately 2.8% [51].

The resolutions of the track transverse and longitudinal impact parameters have been first measured using collisions data in 2010 [71]. For tracks with $p_T \gtrsim 8$ GeV, the transverse impact parameter resolution ranges between 25 μm and 50 μm depending on the track pseudorapidity. For the same tracks, the longitudinal impact parameter resolution increases from 30 μm in the barrel region to 300 μm for $|\eta| \simeq 2.5$.

Vertex reconstruction

Reconstructed tracks are used to determine the primary vertices associated to hadron-hadron collisions. These include the vertices originating from additional collisions (pileup) in the same LHC bunch-crossing.

Vertex reconstruction is performed in two separate stages. First, selected tracks in each event are grouped into clusters, each associated to a separate collision. This clustering procedure is based on a *deterministic annealing* algorithm [72]. The latter consists in an iterative method, which starts from a loose clustering in which different clusters can share

multiple tracks and proceeds to the identification of separate clusters. The algorithm is capable of resolving vertices with a longitudinal separation of approximately 1 mm and it is expected to remain efficient in the high pileup multiplicity environment of LHC collisions. The second part of vertex reconstruction consists in the determination of the vertex properties, in particular its spatial coordinates. An *adaptive vertex fitting* technique [73, 74] is used for this purpose. The algorithm fits the vertex position based on the kinematics of the associated tracks, it rejects outlier tracks and it assigns to each of the remaining tracks a weight proportional to the level of compatibility between the track kinematics and the vertex position.

For vertices reconstructed using at least fifty tracks, the spatial resolution of the vertex position lies between 10 μm and 12 μm for each of the three spatial dimensions [51].

3.3 Muon reconstruction and identification

Muon reconstruction [75] is based on combining the information from the inner tracking system with that of the muon spectrometers. Two complementary reconstruction methods are used in CMS.

- *Global muon reconstruction* (outside-in): a track reconstructed in the muon system (standalone-muon track) is matched to a compatible track found in the inner tracking system (tracker track); a global-muon track is obtained by simultaneously fitting the corresponding hits in the tracker and muon chambers with a method based on the Kalman filter [70].

- *Tracker muon reconstruction* (inside-out): a tracker muon corresponds to a tracker track whose extrapolation to the muon system is compatible with the position of at least one segment in the muon chambers.

More than 99% of the muons produced with sufficient p_T within the geometric acceptance of the muon system ($|\eta| < 2.4$) are reconstructed as either a tracker-muon or a global-muon. The vast majority of muons are reconstructed by both methods and, in this case, muon candidates in the two categories sharing the same tracker track are merged into a single candidate. The muon p_T resolution for muons with $p_T < 100$ GeV ranges from 1% to 6% depending on the muon pseudorapidity. Measurements with muons from cosmic rays show that the p_T resolution stays below 10% for muons with $p_T \lesssim 1$ TeV in the barrel region [75].

After the baseline reconstruction, muon candidates are required to pass a set of reconstruction quality criteria, also called identification (ID) criteria. Different ID methods have been developed in CMS, each leading to different levels of efficiency and purity. The choice of the specific ID method to be used depends, in general, on the physics analysis. The muon candidates used in the analyses described in this work are required to be reconstructed by the Particle-Flow algorithm (PF muons) [69]. Two types of muon ID requirements are of particular relevance for the results presented in later chapters.

- *Tight muon identification*: this ID selection requires that the muon candidate is reconstructed as both a PF-muon and a global-muon, and the normalized-χ^2 (χ^2/NDF) of the global-muon track fit is lower than ten. The corresponding tracker track is also required to contain more than ten hits in the inner tracks (including at least one pixel hit) and to be matched to muon segments in at least two muon stations. Finally, tight cuts are applied on the maximum transverse and longitudinal impact parameters of the muon candidate with respect to the position of the event primary vertex.

- *Medium muon identification*: this identification method was developed for the analysis of CMS data in the LHC Run-2 [76]. Muons passing the Medium-ID working point are required to be PF muons, and either global muons or tracker muons. The method exploits a set of compatibility requirements between the muon tracker track and the associated segments in the muon chambers. Cuts related to the muon impact parameters (and isolation) are not included in this ID. The first results in data and simulation showed that the Medium-ID provides a mis-identification rate comparable to that of the Tight-ID, but a higher efficiency for prompt muons. A comparison of the performance of the two algorithms can be found in Section 5.4.1.

Fixed-cone isolation

An additional requirement used in most physics analyses to select prompt muons relies on the so-called PF fixed-cone isolation. This variable is defined as

$$I^R_{\mathrm{PF}}(\mu) = \frac{\sum p_T^{\text{CH-PV}} + \max\left(\sum p_T^{\mathrm{NH}} + \sum E_T^{\gamma} - \alpha_{\mathrm{PU}},\ 0\right)}{p_T^{\mu}} \tag{3.1}$$

where the three sums in the numerator run over the PF candidates identified as charged hadrons associated to the primary vertex (CH-PV), neutral hadrons (NH) and photons (γ). Each sum considers only the PF candidates found within a cone of fixed-radius R (typically $R = 0.3$ or $R = 0.4$) with respect to the momentum of the muon candidate. The term α_{PU} is a correction used to subtract the contribution of neutral particles from pileup interactions; the specific form of this correction can be either based on the PF algorithm ($\Delta\beta$-correction) or on the average energy density measured in the event (effective-area correction).

In the most conventional event topologies, the value of I^R_{PF} is expected to be small for signal muons produced directly in the hard-scattering process. On the other hand, it is important to stress that this isolation variable is not suited for the identification of muons originating from the decay of high-p_T top quarks. In the latter case, a prompt muon is expected to be emitted with limited angular separation from the decay of a bottom quark; such a muon is thus characterized by a high value of the fixed-cone isolation given in Equation (3.1). Based on this, the analyses described in Chapter 4 and Chapter 5 substitute the conventional cut on isolation with a dedicated method yielding a higher efficiency for boosted top quark decays. A study on simulated events comparing the performance of the latter method and fixed-cone isolation in the context of $X \to t\bar{t}$ searches can be found in Section 5.4.1.3.

3.4 Electron reconstruction and identification

Electron reconstruction [77] is comprised of three basic component: the determination of the energy deposits in the ECAL detector, the reconstruction of the electron track in the silicon tracker and the association of track and ECAL cluster.

Two complementary methods are used to reconstruct electrons. The first method starts from the reconstruction of energy deposits in the ECAL. Adjacent crystals with energy deposits above predefined thresholds are clustered together into ECAL *superclusters* (SC), seeded by the crystal with the highest energy. These superclusters are then utilized to search for a compatible track in the silicon detector. The second method, based on the PF algorithm, is initiated by electron track candidates in the silicon detector, which are extrapolated to the ECAL to find a matching supercluster. This tracker-seeded method is useful to recover efficiency for low-p_T electrons and non-isolated electrons.

Electrons are generally affected by significant energy losses via bremsstrahlung as they cross the tracker layers before reaching the ECAL surface. Each energy loss modifies the curvature of the track trajectory and this requires a dedicated algorithm for the optimal reconstruction of the electron track. This is achieved with a method known as the Gaussian Sum Filter (GSF) algorithm [78], a modified version of the Kalman filter which correctly models the electron energy loss due to photon emissions.

Based on measurements with collisions data at $\sqrt{s} = 8$ TeV [77], the electron reconstruction efficiency for 10 GeV $< p_T^e <$ 100 GeV ranges from 88% to 98%, depending on the electron pseudorapidity. In a similar p_T range, the electron energy resolution varies between 1.7% and 4.5%, depending on the electron pseudorapidity. Due to the high resolution of the ECAL energy measurements, the energy resolution improves for high-p_T electrons, reaching a value of approximately 0.5% for $p_T^e > 100$ GeV.

Electron identification methods are used to distinguish prompt electrons from background sources, mainly originating from photon conversions and jets mis-identified as electrons. These methods rely on a set of reconstruction quantities, which can be classified as follows.

- *Track-cluster compatibility*: the compatibility between measurements in the ECAL and in the tracker is quantified using geometric criteria and comparing the supercluster energy to the track momentum.

- *Cluster properties*: calorimetric quantities based on the transverse shape of the ECAL showers are used to discriminate true electrons (prompt electrons or electrons from photon conversions) from mis-identified electrons (e.g. jets with large electromagnetic components), exploiting the fact that electromagnetic showers are narrower than hadronic showers. The energy fractions deposited in the HCAL, which are expected to be small for electromagnetic showers, are also used.

- *Track properties*: tracking variables are employed to improve the distinction between electrons and charged hadrons, comparing the results of the GSF-fitted track to the prediction of the standard Kalman filter.

- *Rejection of photon conversions*: electrons from photon conversions are identified from tracks with missing hits in the inner layers of the silicon tracker and track pairs with opposite electric charges originating from a common vertex and having similar tangents at the conversion vertex, as expected for the kinematics of $\gamma \to e^+e^-$ decays.

These variables are generally used in two ways for the purposes of electron ID: one possibility is to design an identification method implemented with a cut-based approach, while a second option is to perform a multivariate analysis (MVA) and determine a single discriminator providing the best separation between signal and background electrons.

Similarly to the case of muons, electron identification methods can be separated from the isolation requirement. The discussion on PF-based fixed-cone isolation given in Section 3.3 can be entirely extended to electrons; this includes the considerations on the ineffectiveness of such a variable to select electrons from boosted top quark decays in the context of $X \to t\bar{t}$ searches. As for muons, the electron candidates used in the analyses of Chapter 4 and Chapter 5 are not required to pass a cut on fixed-cone isolation.

3.5 Jets and missing transverse energy

3.5.1 Jet reconstruction

The hadronization of quarks (or gluons) produced in the LHC collisions gives rise to showers of stable particles (hadrons, photons and leptons) in the detector. These particles, once reconstructed, are clustered into so-called *jets* in order to determine the kinematics of the original parton. In general, jet clustering can be performed using the energy deposits in the calorimeters (Calo-jets) or the list of particle candidates provided by the PF algorithm (PF-jets). The PF-based reconstruction has been shown to significantly improve the measurement of the jet energy and direction compared to the calorimeter-based approach [66]. For this reason, the analyses described in this work rely exclusively on the use of PF-jets in the offline reconstruction.

The candidates reconstructed by the PF algorithm are clustered according to the following procedure. The clustering starts by identifying the values of the distances d_{ij} between two candidates (i and j) and the distance d_i assigned to each single candidate. If the smallest value corresponds to a d_{ij} quantity, the two candidates are combined by summing their four-momenta; on the other hand, if the smallest value is given by a d_i quantity, the candidate i is identified as a jet and removed from the clustering inputs. After each iteration, these distances are recalculated and the procedure is repeated until no candidates are left. The distances d_{ij} and d_i are defined as

$$d_{ij} = \min\left(p_{Ti}^{2k}, p_{Tj}^{2k}\right) \cdot \frac{\Delta_{ij}^2}{R^2} \qquad \text{and} \qquad d_i = p_{Ti}^{2k} \tag{3.2}$$

where $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ and p_{Ti}, y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of candidate i, respectively. The radius parameter R is a constant. The value of the integer k defines the type of jet-clustering algorithm. The choice $k = 0$ identifies the Cambridge-Aachen algorithm [79], which is used in specific methods designed to identify hadronic top quark decays, as detailed in Section 3.5.4. The case of $k = -1$ corresponds to the anti-k_T algorithm [80]. The latter is usually the preferred method for jet clustering in LHC analyses. This is motivated by the fact that jets reconstructed with the anti-k_T algorithm are resilient against the emission of soft radiation (infrared safety) and the presence of collinear decay products from particle splitting (collinear safety); in addition, anti-k_T jets have a simple and intuitive cone structure in the η-ϕ plane.

PF candidates identified as charged hadrons with an inner track not associated to the event primary vertex are not considered in the clustering sequence. This approach, known as *Charged Hadron Subtraction* (CHS) [81], reduces the energy contribution of charged particles from pileup interactions.

The jet energy scale is corrected using different calibration methods. Several types of jet energy corrections (JEC) are applied in a modular approach in order to account for the contamination from neutral pileup particles and detector noise ("L1-Offset") and the non-linearity of the calorimeter response in the jet η ("L2-Relative") and jet p_T ("L3-Absolute"); an additional correction is applied only in data events to account for residual differences between the jet response in data and simulation ("L2L3-Residual"). These correction factors are determined as a function of the jet p_T and η and they are used to rescale the four-momentum of the original jet. Based on the latest measurements with pp collisions data at $\sqrt{s} = 8$ TeV [82], the overall uncertainty on the jet energy scale for $p_T > 30$ GeV is lower than 1% in the barrel region ($|\eta| < 1.3$) and it remains below 3% for $|\eta| < 5.0$.

The typical jet energy resolution (JER) for jets in the barrel region is approximately 15%-20% for $p_T \simeq 30$ GeV, 10% for $p_T \simeq 100$ GeV and 5% for $p_T \simeq 1$ TeV [82]. The jet energy resolutions are measured in data and MC events and the resulting data/MC scale factors are used to correct the JER in simulated events.

The collection of PF candidates used in the jet clustering potentially includes the particles which are identified as muons and electron candidates at a later stage of the analysis. In order to solve this ambiguity and avoid any double-counting of PF candidates in the reconstruction of the event, a procedure generally referred to as *jet-lepton cleaning* is required. The latter can be implemented in different ways depending on the analysis of interest. The methods used in each of the analyses described in this work are detailed in the corresponding chapters.

3.5.2 Missing E_T reconstruction

In a hadron-hadron collision, the longitudinal momentum of the scattering process between the initial-state protons cannot be measured, while the the total momentum in the plane transverse to the beam direction is expected to be negligible. Based on this, the measurement of the imbalance of the transverse energy in the final state (missing E_T, $\not\!\!E_T$) can be used to infer the production of particles that cannot be directly detected by the experimental apparatus. Examples of such particles are the neutrinos of the SM and other non-interacting particles predicted in many BSM scenarios.

Similarly to the case of jets, the optimal method for the measurement of the missing E_T in CMS is based on the PF algorithm [66]. The uncorrected $\not\!\!E_T$ vector is given by the negative vectorial sum of the transverse momenta of all the PF candidates in the event:

$$\vec{\not\!\!E}_T^{\,\text{unc}} = -\sum_{i \notin \text{jets}} \vec{p}_{T,\,i} - \sum_{i \in \text{jets}} \vec{p}_{T,\,i} = -\sum_{i \notin \text{jets}} \vec{p}_{T,\,i} - \sum_{\text{jets}} \vec{p}^{\,\text{unc}}_{T,\,\text{jet}} \,. \tag{3.3}$$

where the first term on the right-hand side corresponds to the contribution from candidates not clustered in jets and the second term denotes the contribution from candidates included in the jet-clustering; the latter term corresponds to the vector sum over the uncorrected $\vec{p}_T$ of the jets. The measurement of $\not\!\!E_T$ can be corrected to include the effect of the jet energy corrections. This improved $\not\!\!E_T$ estimate, referred to as "Type-1 PF-MET", is given by

$$\vec{\not\!\!E}_T^{\,\text{Type-1}} = \vec{\not\!\!E}_T^{\,\text{unc}} + \vec{C}_T^{\,\text{Type-1}} \tag{3.4}$$

where the term $\vec{C}_T^{\,\text{Type-1}}$, known as Type-1 correction, corresponds to

$$\vec{C}_T^{\,\text{Type-1}} = \sum_{\substack{\text{jet} \\ p^{\text{corr}}_{T,\,\text{jet}} > 10 \text{ GeV}}} \left(\vec{p}^{\,\text{unc}}_{T,\,\text{jet}} - \vec{p}^{\,\text{corr}}_{T,\,\text{jet}}\right) \tag{3.5}$$

This correction accounts for the propagation of the JECs to the $\not\!\!E_T$ measurement, in a phase space where such corrections can be measured with sufficient accuracy ($p^{\text{corr}}_{T,\,\text{jet}} > 10$ GeV).

Even in the absence of undetected particles inducing true missing energy, the value of $\not\!\!E_T$ can differ from zero because of several effects. Some of them are related to the experimental apparatus (e.g. detector noise, detector energy resolution, mis-reconstruction of particles and jet energy corrections), while others are due to other physics processes, such as pileup interactions and underlying event activity. The impact of these effects and the performance of the PF $\not\!\!E_T$ reconstruction, in particular its energy scale and resolution, have been measured with pp collisions data in $\text{Z} \to e^+e^-$, $\text{Z} \to \mu^+\mu^-$ and single-photon events [83].

3.5.3 Identification of jets from bottom quark decays

The identification of jets originating from the decay of bottom quarks, i.e. jet b-tagging, is a very powerful tool to isolate physics processes characterized by b-quark production, e.g. $t\bar{t}$ production and $h \to b\bar{b}$ decays. Jet b-tagging algorithms exploit the particular kinematics induced by the relatively long lifetime of B-hadrons, given by tracks with large impact parameter with respect to the primary vertex and possibly a secondary vertex reconstructed inside the jet, as well as the presence of soft leptons from semileptonic B-hadron decays.

Several methods have been developed in the CMS experiment for the purpose of jet b-tagging [84]. The algorithm utilized in the analyses presented in this thesis is the *Combined Secondary Vertex* (CSV) algorithm. The latter method utilizes multiple reconstruction quantities related to the tracks associated to the jet constituents and the properties of secondary vertices reconstructed inside the jet, if present. The CSV algorithm is still applicable when no secondary vertices are associated to the jet, so its performance is not inherently limited by the secondary vertex reconstruction efficiency. The variables used by this method include the number of tracks in the jet, the significance of the tracks' impact parameters, and, if available, the impact parameter, mass and number of tracks associated to the secondary vertex. These variables are used to determine a single likelihood discriminator associated to the jet, i.e. the CSV discriminator, proportional to the compatibility of the jet with a b-quark decay.

Three working points of the CSV algorithm, named "loose" (CSVL), "medium" (CSVM) and "tight" (CSVT), are defined based on their b-tagging mis-identification (or mistag) rate, i.e. the probability to b-tag a light-flavor jet, which is equal to 10%, 1% and 0.1%, respectively. For these same working points, the b-tagging efficiency for heavy-flavor jets is typically around 85% (CSVL), 65% (CSVM) and 45% (CSVT).

The performance of the CSV algorithm (and others) has been validated extensively using pp collisions data in the LHC Run-1 [84,85] and Run-2 [86]. The b-tagging mis-identification rate is measured in data using a sample of dijet events dominated by QCD-multijet production. The b-tagging efficiency for heavy-flavor jets is measured in independent samples of QCD events enhanced in B-hadron production (identified using the B-hadron semileptonic decays) and $t\bar{t}$ events in the ℓ + jets and dilepton final states.

3.5.4 Identification of merged jets from top quark decays

The unprecedented c.o.m. energy of the pp collisions produced at the LHC has opened the possibility to directly probe the existence of new particles with a mass above the TeV scale. The decay products of such a particle are expected to be highly energetic; in the case where one of them is a top quark decaying hadronically, the latter can produce very collimated decay products which can be reconstructed in the detector as a single merged jet with specific substructure properties. More precisely, for a top quark with a Lorentz boost $\gamma = E/m$, the angular separation between the W boson and b quark from the top quark decay will be of the order of $\Delta R = 2/\gamma$. Therefore, if the W boson decays hadronically, the products of the W boson and b quark decays, which typically form three separated jets, could be contained in a single jet clustered with a distance parameter R larger than $2/\gamma$, with the latter jet characterized by an energy distribution compatible with the decay of three quarks.

The methods aimed at distinguish such merged jets from normal jets from light-quark (or gluon) decays are known as t-tagging algorithms [34, 35]. These techniques have gained increasing relevance in the analysis of LHC data, not only in BSM searches involving top quarks, but also in dedicated SM measurements of top quark production in the boosted

regime. Several algorithms have been developed in recent years to optimize the identification of merged jets from boosted top quark decays at the LHC. Many of these methods rely on a general approach known as *jet grooming*. The latter consists in the removal of soft and wide-angle radiation from the jet, in order to mitigate the contamination due to initial state radiation, underlying event and pileup interactions. One of the main applications of this approach is the determination of the groomed jet mass, a new mass estimator which can provide better discrimination, compared to the plain jet mass, between true top quark jets and jets seeded by light partons.

In the following, we briefly review some of the methods used in the CMS experiment for jet t-tagging in the LHC Run-1 and Run-2 [87, 88]. In particular, we focus on the t-tagging algorithms employed in the physics analyses of Chapter 4 and Chapter 5.

CMS Top Tagging algorithm

The CMS Top Tagging (CMSTT) algorithm [34, 89], relies on the direct decomposition of the input jet in order to determine its subjets and calculate variables suited to distinguish top quark jets from QCD jets.

The input of the CMSTT algorithm are Cambridge-Aachen (CA) jets clustered with a distance parameter $R = 0.8$. The grooming method consists in two iterations of the same jet decomposition procedure: the first one (primary decomposition) is applied on the original CA8 jet; the second one (secondary decomposition), if possible, is applied on each of the subclusters found in the first iteration. The decomposition algorithm proceeds as follows.

1. The last step of the jet clustering procedure is inverted to find two jet subclusters.

2. The method proceeds to the following step if the two subclusters satisfy

$$\Delta R > 0.4 - A \cdot p_T^C \,, \tag{3.6}$$

   where p_T^C is the transverse momentum of the original jet in the primary decomposition (or a subcluster in the secondary decomposition) and $A = 4 \cdot 10^{-4}$ GeV^{-1} is a slope parameter tuned using simulated events. If this selection fails, the decomposition fails.

3. The decomposition succeeds if both subclusters satisfy the p_T fraction criterion

$$p_T^{\text{cluster}} > \delta_P \cdot p_T^{\text{jet}} \,, \tag{3.7}$$

   with $\delta_P = 0.05$. The latter value was tuned in simulation studies in order to minimize the mistag rate for a given signal efficiency. If only one of the subclusters meets the requirement in Equation (3.7), the decomposition procedure is repeated only on that subcluster. If, after this iterative process, there is no subcluster satisfying Equation (3.7), or the subcluster is a single constituent, the decomposition fails.

Based on the above procedure, each jet can be decomposed into two, three or four subjets. The sum of the four-momenta of the reconstructed CMSTT subjets corresponds to the four-momentum of the groomed jet. The final t-tagging selection relies on a cut-based approach involving the following quantities:

- jet mass (m_{jet}): the invariant mass of the original CA8 jet; top quark jets are expected to have a mass close to the top quark mass.

- number of subjets (N_{subjets}): the number of subjets determined by the CMSTT algorithm; top quark jets are expected to have at least three subjets.

- minimum subjet pairwise mass (m_{min}): the minimum value of the invariant masses of all possible pairs involving the three highest-p_T subjets, i.e. $m_{\text{min}} \equiv \min(m_{12}, m_{13}, m_{23})$; this quantity is expected to be close to the W boson mass for a merged jet associated to a hadronic top quark decay; in the case $N_{\text{subjets}} = 2$, this variable is not defined.

The specific thresholds on m_{jet} and m_{min} can be tuned in order to reach the desired t-tagging efficiency or mistag rate. The CMSTT algorithm is at the basis of the jet t-tagging method used in the physics analysis described in Chapter 4.

Softdrop declustering

Another important example of grooming technique is the *softdrop declustering* algorithm [90]. The method proceeds as follows: given a jet clustered with a distance parameter R, the jet clustering sequence is inverted step by step, starting from the original jet. The starting jet (or pseudo-jet) j is split into two subjets j_1 and j_2 and, if the softdrop condition

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta} \tag{3.8}$$

is satisfied, the declustering stops; otherwise, the subjet with the lower p_T is discarded and the declustering is repeated on the higher-p_T subjet. The invariant mass of the jet constituents not rejected by this grooming procedure is referred to as the jet *softdrop mass*. Compared to the mass of the ungroomed jet, the softdrop mass is less dependent on the jet p_T, has a better resolution and a mean value closer to the mass of the jet-seeding particle.

The softdrop algorithm contains two free parameters, z_{cut} and β, which control the strength of the jet grooming procedure; in the limit $\beta \to \infty$ or $z_{cut} \to 0$, no grooming is applied to the original jet. A typical choice for the two parameters is $z_{cut} = 0.1$ and $\beta = 0$. The softdrop algorithm is employed in the t-tagging selection applied in the analysis presented in Chapter 5.

N-subjettiness

The so-called N-subjettiness variable [91] is a jet-shape observable designed to quantify how consistent a jet is with the hypothesis of being composed of N subjets. This quantity, commonly denoted as τ_N, is defined as follows

$$\tau_N = \frac{\sum_i^{n_c} p_{Ti} \cdot \min\left(\Delta R_{i,1}, \ldots, \Delta R_{i,N}\right)}{\sum_i^{n_c} p_{Ti} \cdot R} \tag{3.9}$$

where both sums run over the n_c constituents of the jet, R is the parameter used in the original jet-clustering sequence and p_{Ti} is the transverse momentum of the i-th jet constituent. The symbol $\Delta R_{i,K}$ ($K = 1, \ldots, N$) denotes the ΔR distance between the i-th jet constituent and the axis of the K-th subjet. The N subjet axes used in the numerator of Equation (3.9) are the ones minimizing the value of τ_N and they are determined with a procedure seeded by the reclustering of the jet constituents with the k_T algorithm [92]. The quantity τ_N thus corresponds to the p_T-weighted average of the angular distances between the jet constituents and their closest subjet. If a jet is consistent with having N subjets, the value of τ_N will tend to zero; otherwise, τ_N will be close to unity.

The ratio between τ_3 and τ_2, i.e. $\tau_{32} \equiv \tau_3/\tau_2$, is found to be a powerful discriminant between jets originating from hadronic top quark decays and jets from gluon and single-quark hadronization. This quantity is used for the purpose of jet t-tagging in both physics analyses presented in this thesis.

Chapter 4

Search for $t\bar{t}$ resonances in the ℓ + jets channel at $\sqrt{s} = 8$ TeV

This chapter presents a search for a resonance decaying to a $t\bar{t}$ pair in semileptonic final states (μ+jets and e+jets). The analysis is based on the full integrated luminosity recorded by the CMS experiment in pp collisions at a center-of-mass energy of $\sqrt{s} = 8$ TeV. At the time this data was collected (2012), this represented the highest energy reached for hadron-hadron collisions in a particle physics experiment.

The results of this analysis are part of a publication by the CMS Collaboration on the combination of searches for resonant $t\bar{t}$ production at $\sqrt{s} = 8$ TeV [3].

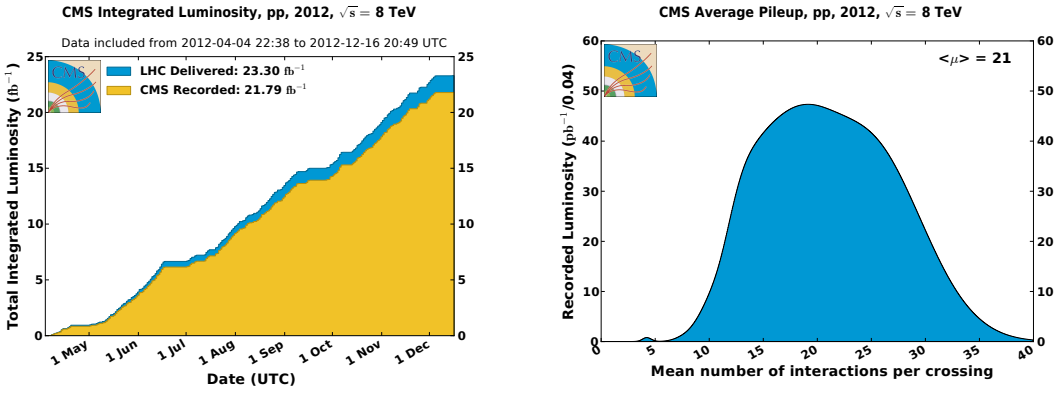
4.1 Data-taking operations in 2012

Stable proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 8$ TeV were produced at the LHC for the first time in April 2012. Throughout the course of 2012, the LHC delivered a total integrated luminosity of 23.3 fb^{-1} to both the ATLAS and CMS experiments, reaching a peak instantaneous luminosity of around $7.7 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. This level of instantaneous luminosity was obtained by colliding proton bunches with a 50 ns bunch spacing.

The CMS experiment recorded a total integrated luminosity equal to 21.8 fb^{-1}, for an overall data-taking efficiency of 93.5%; the integrated luminosity recorded by CMS as a function of time in 2012 is shown in Figure 4.1(a). This entire data set was recorded with a fully operational magnetic field of $B = 3.8$ T. The average number of pileup interactions measured in CMS during the 2012 data-taking campaign was approximately $\langle\mu\rangle = 21$. Figure 4.1(b) shows the full luminosity profile of the $\sqrt{s} = 8$ TeV data set as a function of the average number of pileup collisions.

4.2 Data and Monte Carlo samples

This analysis is based on data collected by the CMS experiment during the 2012 LHC Run with a fully operational detector and a magnetic field of $B = 3.8\,\mathrm{T}$. This corresponds to a data sample with a total integrated luminosity of 19.7 fb^{-1}. We utilize the data sets produced in the most recent re-processing of the 2012 CMS data set, which include the latest corrections for the alignment and calibration of the CMS subdetectors. These samples have been produced centrally using the official reconstruction software of the CMS experiment (CMSSW, release 5.3). The data stream of events recorded with muon-based triggers is used



(a) (b)

Figure 4.1: (a) Total integrated luminosity delivered to and recorded by the CMS experiment in 2012 in pp collisions at $\sqrt{s} = 8$ TeV. (b) Integrated luminosity in the CMS pp data set of 2012, as a function of the average number of pileup interactions.

for the analysis in μ + jets channel. In the e + jets analysis, the HLT selection is given by the combination of two separate triggers, an electron-based one and a jet-based one; as a result, the electron channel includes events coming from two different data streams (the `ElectronHad` data stream and the `JetHT` data stream). A detailed description of the trigger strategy employed in the analysis can be found in Section 4.5.1.

Monte Carlo (MC) simulated events are used to study the properties of background and signal processes. All simulated events are generated with a center-of-mass energy of $\sqrt{s} = 8$ TeV and they are reconstructed with the same software used to reconstruct data events. Every MC sample includes the simulation of additional inelastic proton-proton interactions, including both in-time and out-of-time pileup. The original pileup distribution in the simulation was designed to cover a wide range of pileup multiplicities, in order to allow for a reweighting of the MC sample to the actual pileup conditions determined after data-taking. For all MC samples, unless otherwise specified, the parton shower and hadronization processes are simulated with the PYTHIA-v6.4 generator [93]; the same generator is also used to model the underlying event activity using the `TuneZ2Star` tune [94]. The response of the CMS subdetectors is simulated using the GEANT4 package [95]. In the remainder of this section, we detail the specifics of the simulated samples used for the SM and BSM processes relevant to the analysis.

4.2.1 SM backgrounds

The production of a top-antitop quark pair in the SM is simulated with the next-to-leading order (NLO) generator `POWHEG`-v1 [96,97,98]. Two exclusive samples, binned with respect to the mass of the generated $t\bar{t}$ pair (0.7 TeV $< M_{t\bar{t}}^{\text{GEN}} <$ 1 TeV and $M_{t\bar{t}}^{\text{GEN}} >$ 1 TeV), are used in combination with an inclusive sample in order to increase the MC statistics in the high-$M_{t\bar{t}}$ region. In each of these MC samples, all $t\bar{t}$ decay modes are included. The prediction of the $t\bar{t}$ MC is normalized to the inclusive $t\bar{t}$ cross section calculated at NNLO in QCD [99]. The LO generator `MadGraph`-v5 [100] is used for the simulation of W and Z boson production in association with jets. The MC matrix element includes up to four additional partons and the matching with the parton shower simulation is performed according to the MLM matching scheme [101]. The inclusive cross sections for W and Z production calculated up to QCD-

NNLO accuracy with FEWZ-v3.1 [102] are used to normalize the MC prediction. Single-top production in the s-channel, t-channel and tW-channel is simulated with POWHEG-v1; the inclusive cross sections calculated up to approximate QCD-NNLO accuracy in Ref. [33] are used to normalize these MC samples. Diboson production (WW, WZ, ZZ) is simulated at LO with PYTHIA-v6.4; the inclusive cross sections for these three processes are calculated at NLO in QCD using MCFM-v6.6 [103].

The background MC samples used in the analysis make use of different sets of parton-distribution-functions (PDF): the NLO CT10 PDF set [104] is used in the simulation of SM $t\bar{t}$ production, while all the other background samples are generated using the CTEQ6 PDF parameterization [105, 106].

Table 4.1 lists the MC samples used for the simulation of SM processes and the corresponding cross sections.

Table 4.1: MC samples and cross sections used in the analysis for SM processes.

process	samples	MC generator	cross section [pb]
$t\bar{t}$ + jets	$t\bar{t}$ (inclusive)	POWHEG-v1	245.8
	$t\bar{t}$ (0.7 TeV $< M_{t\bar{t}}^{\mathrm{GEN}} <$ 1 TeV)	POWHEG-v1	18.2
	$t\bar{t}$ ($M_{t\bar{t}}^{\mathrm{GEN}} >$ 1 TeV)	POWHEG-v1	3.4
W + jets	W($\to \ell\nu$) + 1 jets	MadGraph-v5	6663
	W($\to \ell\nu$) + 2 jets	MadGraph-v5	2159
	W($\to \ell\nu$) + 3 jets	MadGraph-v5	640
	W($\to \ell\nu$)+ $\geqslant$ 4 jets	MadGraph-v5	264
Z + jets	Z($\to \ell\ell$) + 1 jets ($M_{\ell\ell}^{\mathrm{GEN}} >$ 50 GeV)	MadGraph-v5	666
	Z($\to \ell\ell$) + 2 jets ($M_{\ell\ell}^{\mathrm{GEN}} >$ 50 GeV)	MadGraph-v5	215
	Z($\to \ell\ell$) + 3 jets ($M_{\ell\ell}^{\mathrm{GEN}} >$ 50 GeV)	MadGraph-v5	60.7
	Z($\to \ell\ell$)+ $\geqslant$ 4 jets ($M_{\ell\ell}^{\mathrm{GEN}} >$ 50 GeV)	MadGraph-v5	27.4
single top	s-channel	POWHEG-v1	5.6
	t-channel	POWHEG-v1	87.1
	tW-channel	POWHEG-v1	22.2
diboson	WW	PYTHIA-v6.4	54.8
	WZ	PYTHIA-v6.4	33.2
	ZZ	PYTHIA-v6.4	8.1

4.2.2 Signal models

Three types of signal models are considered in the analysis: a color-singlet Z' boson with a relative decay width ($\Gamma_{Z'}/M_{Z'}$) of 1% and 10%, and a Kaluza-Klein excitation of a gluon in the Randall-Sundrum model [107].

The `MadGraph`-v4 generator [108] is used to simulate the production of Z' resonances decaying to a $t\bar{t}$ pair. Up to three extra partons are included in the LO matrix element of this BSM model and the MLM matching algorithm [101] is used to combine the ME calculation with the parton shower simulation. Since the Z' boson is a color-singlet, it is produced only via $q\bar{q}$ annihilation (at LO): this production mechanism does not interference with SM $q\bar{q} \to t\bar{t}$ production at LO and higher-order interference effects are not included in the simulation. The Z' couplings to SM particles are set equal to those of the SM Z boson: this implies that the resonance has a larger coupling with left-handed top quarks compared to right-handed top quarks. Polarization effects in top quark decays are included in the $Z' \to t\bar{t}$ simulation using the `DECAY` plugin of `MadGraph`-v4.

The KK gluon signal samples are generated with PYTHIA-v8.1 [109] using the implementation described in Ref. [110]. Like in the case of the Z' model, this massive KK resonance is produced only via $q\bar{q}$ annihilation at LO [111]. Since the KK gluon resonance is a color-octet, its production interferes with SM $q\bar{q} \to t\bar{t}$ production already at LO, but interference effects are not included in the MC samples used for this analysis. One peculiarity of the KK gluon resonance in the RS model is that its coupling to right-handed top quarks is approximately five times bigger than the corresponding SM coupling ($gt\bar{t}$ vertex), whereas the coupling to left-handed top quarks is the same as its SM counterpart. Top quarks originating from this BSM particle are thus expected to be highly polarized. On the other hand, the event generator used for these signal samples (PYTHIA-v8.1) does not include the full treatment of polarization effects and spin correlations for the generated top quarks: as a result, the top quarks produced in the decay of the KK gluon resonance are effectively treated as unpolarized.

Simulated samples for the Z' (KK gluon) signals are generated for several mass hypotheses between 0.5 TeV (0.7 TeV) and 4 TeV. In each of these signal samples, the resonance decays to a $t\bar{t}$ pair and each top quark can decay in any of its decay modes, according to its SM branching fractions. All signal samples are generated using the LO CTEQ6 PDF set.

For the $Z' \to t\bar{t}$ signals, we use as reference the cross sections calculated at LO for a leptophobic Z' in the topcolor model [112]. The reference cross sections for the production of a KK gluon in the RS model are also calculated at LO [24]. These cross sections are multiplied by a factor $K = 1.3$ in order to account for higher-order corrections [113] and they are shown in Figure 4.2. The invariant mass distribution of the generated top-antitop quark pair, before any event selection, can be seen in Figure 4.3 for each of the three signal models. These distributions show that the fraction of events characterized by the off-shell production of the BSM resonance is proportional to the width of the resonance. This effect arises from the convolution of the parton-level cross section with the proton parton distribution functions; the effect is particularly important for the 10%-width Z' and the KK gluon resonance, which is characterized by an even larger relative width (between 15% and 20%).

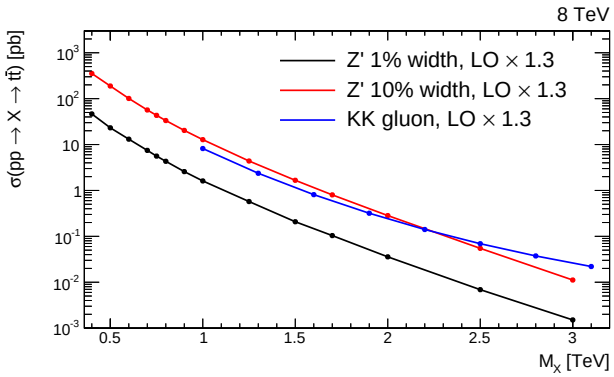


Figure 4.2: Cross sections times branching ratio to $t\bar{t}$, i.e. $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$, for the signal models considered in the analysis. The cross sections of a leptophobic $Z' \to t\bar{t}$ signal are calculated at LO in the topcolor model [112]. The cross sections for the production of a KK gluon resonance in the RS model are also determined up to LO accuracy [24]. These cross sections are multiplied by a factor $K = 1.3$ to account for higher-order corrections [113].

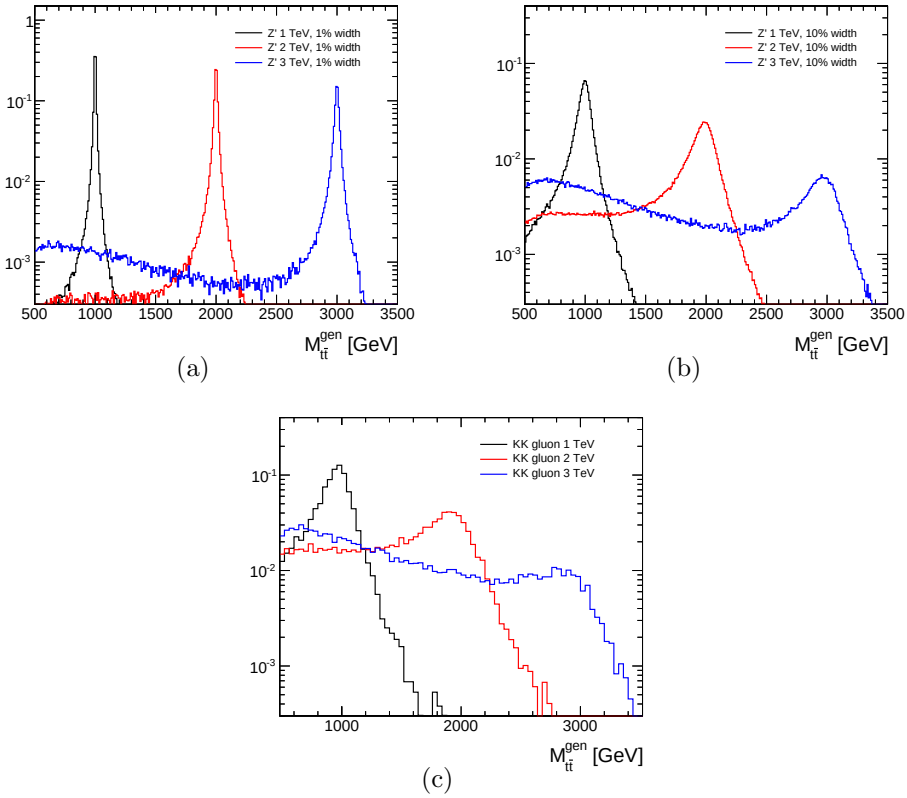


Figure 4.3: Invariant mass distribution of the generated $t\bar{t}$ pair, before any event selection, for each of the three signal models considered in the analysis: (a) color-singlet Z' boson with 1% relative decay width, (b) color-singlet Z' boson with 10% relative decay width, (c) KK excitation of a gluon in the RS model. In each plot, distributions are shown for signal mass hypotheses of 1 TeV, 2 TeV and 3 TeV.

4.3 Object reconstruction

The event reconstruction relies on the global event description provided by the CMS Particle Flow algorithm [66] (see Section 3.1). This section describes the specific requirements used to identify the primary vertex of the event and the physics object candidates used in the analysis, which include muons, electrons, hadronic jets and missing transverse energy.

4.3.1 Primary vertices

Primary vertices are determined from the tracks reconstructed in the event, according to the methods described in Section 3.2. Each vertex candidate is required to satisfy a set of quality criteria: $\sqrt{x^2 + y^2} < 2$ cm, $|z| < 24$ cm, where x, y and z denote the coordinates of the vertex position, and $N_{\text{DOF}} > 4$, where N_{DOF} is the weighted number of tracks associated to the vertex. The vertex candidate with the highest $\sum_{\text{tracks}} p_T^2$ is identified as the primary vertex of the event.

Pileup reweighting for MC samples

In order to reproduce in the simulation the average number of interactions measured during the 2012 data-taking, simulated samples are reweighted on an event-by-event basis. The pileup correction is implemented as follows. For collisions data, the number of interactions in the event is obtained by multiplying the instantaneous luminosity per bunch crossing by the total inelastic (minimum-bias) cross section. In MC events, this value corresponds to the true number of interactions simulated in the event. The data/MC ratio of these two distributions yields the weight to be applied on simulated events for a given multiplicity of MC pileup interactions. A value of 69.4 mb is used for the total inelastic cross section and an uncertainty of $\pm 5\%$ is assigned to this cross section estimate to determine the systematic uncertainty on the correction factors.

A closure test of the pileup reweighting method is carried out by looking at the number of primary vertices reconstructed in data and MC events. Figure 4.4 shows this data/MC comparison in a μ + jets sample with a loose selection (one muon with $p_T > 45$ GeV and $|\eta| < 2.1$ and two AK5 jets with $p_T > 50$ GeV and $|\eta| < 2.4$); in the second plot, the reweighted distribution of the number of primary vertices in MC is shown for three different values of the minimum-bias cross section.

4.3.2 Muons

Muon candidates are required to pass tight identification requirements, designed to suppress the contribution of muons from decays in flight and mis-identified hadrons punching through the calorimeters [75]. The muon identification selection corresponds to the following set of criteria, also known as the muon Tight-ID:

- the muon candidate is reconstructed as both a global-muon and a PF-muon;
- the normalized χ^2 of the global-muon track fit passes $\chi^2/N_{\text{DOF}} < 10$;
- there is at least one muon chamber hit included in the global-muon track fit;
- at least two muon stations contain segments matched to the global-muon track;
- $|\Delta z| < 0.5$ cm, where $|\Delta z|$ is the longitudinal distance of the muon inner track with respect to the primary vertex position;

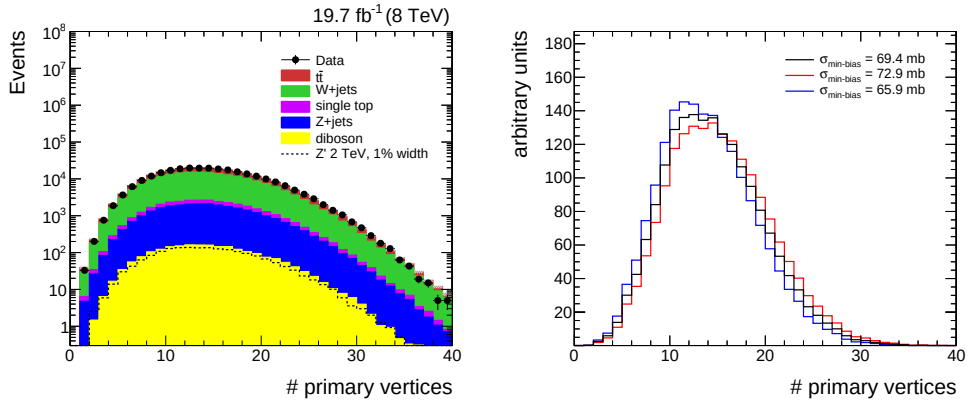


Figure 4.4: Left: data/MC distribution of the number of reconstructed primary vertices, after applying the pileup correction to the simulation, for events with one muon with $p_T > 45$ GeV and $|\eta| < 2.1$, and two AK5 jets with $p_T > 50$ GeV and $|\eta| < 2.4$. The sum of the MC background templates is normalized to data. The background error band includes only the MC statistical uncertainty. Right: comparison between the reweighted MC distribution of the number of primary vertices, using three different values for the minimum-bias cross section: (black) $\sigma_{\text{min-bias}} = 69.4$ mb, (red) $\sigma_{\text{min-bias}} = 72.9$ mb, (blue) $\sigma_{\text{min-bias}} = 65.9$ mb.

- $|d_{xy}| < 0.2$ cm, where $|d_{xy}|$ is the transverse impact parameter of the muon inner track with respect to the primary vertex position;
- there is at least one hit in the pixel detector assigned to the muon inner track;
- more than five tracker layers contain hits associated to the muon inner track.

Each of these muons is required to have $p_T^{\mu} > 45$ GeV and $|\eta^{\mu}| < 2.1$; the relatively high p_T-threshold reduces the contribution of muons coming hadron decays. These are the muon candidates considered in the analysis. In order to account for differences in data and MC due to the muon ID efficiencies, simulated events are corrected using data/MC scale factors measured in a sample of $Z \to \mu^+\mu^-$ decays with the tag-and-probe method [75]; these scale factors are determined as a function of the muon p_T and η. We note that no isolation requirement is applied on the muon candidates, since prompt muons from the decay of high-p_T top quarks are expected to be poorly isolated, due to their proximity to the b-quark decay.

4.3.3 Electrons

Electron candidates are required to satisfy an identification method based on a multivariate (MVA) analysis; in the latter, multiple variables characterizing the quality of the electron reconstruction are combined in a single MVA discriminator in order to achieve a better discrimination power in the identification of prompt electrons [77]. The electron ID requirements include

- a veto on electrons originating from photon conversions,
- at least one hit associated to the electron track in the first tracker layer, and
- a lower cut on the electron MVA discriminator.

Electron candidates used in the analysis are required to have $p_T^e > 35$ GeV and $|\eta_{\rm SC}| < 2.5$, where $\eta_{\rm SC}$ is the pseudorapidity of the electron supercluster in the calorimeter. The relatively tight cut on the electron p_T reduces the contribution of non-prompt electrons. Similarly to muons, the identification algorithm does not include any requirement on the electron isolation. Data/MC scale factors for the electron MVA-based ID are measured as a function of the electron p_T and $\eta_{\rm SC}$ using the tag-and-probe technique in $Z \to e^+e^-$ events [77] and these correction factors are used to reweight the MC samples in the ℓ + jets analysis.

4.3.4 Hadronic jets

Hadronic jets are reconstructed by clustering the PF candidates in the event. Two separate jet collections are considered in the analysis: one contains jets clustered with the anti-k_T (AK) algorithm [80] with cone-parameter $R = 0.5$; the other includes jets clustered with the Cambridge-Aacheen (CA) algorithm with a larger cone radius $R = 0.8$. Jets in the AK5 collection are used to reconstruct the decays of light-flavor (light quarks and gluons) and heavy-flavor (charm, bottom) partons produced in the hard-scattering collision; conversely, larger-radius jets are employed in the analysis to identify clusters associated to the hadronic decay of high-p_T top quarks. For both jet collections, the PF candidates associated to tracks coming from pileup vertices are not included in the jet clustering sequence; this method, known as Charged Hadron Subtraction, allows to minimize the contamination due to charged particles coming from in-time pileup interactions. Every jet candidate in both collections is required to pass minimal jet quality criteria [114].

Jet energy-scale corrections (JEC) are applied to each jet candidate, in both data and simulation, to correct for the energy contribution of neutral particles from pileup and the η- and p_T-dependence of the detector response [115]; an additional set of JECs is applied only in data events to account for residual differences between data and simulation. The same methods are used to derive different sets of energy-scale corrections for the AK5 and CA8 collections. Jet energy-scale resolution (JER) corrections are also applied on jet candidates in MC events, as the jet p_T-resolution measured in data is known to be slightly worse than the one obtained from simulation [116].

Since the PF candidates ultimately identified as leptons (muons or electrons) are initially used in the jet clustering procedure, these candidates need to be removed from the two jet collections; this procedure is usually referred to as *jet-lepton cleaning*. For small-radius jets, a specific jet-lepton cleaning procedure is needed. Due to the characteristic signature of the leptonic decay of a boosted top quark, we do not require a minimum ΔR distance between an AK5 jet and the lepton candidates defined above. If one of the lepton candidates (muon or electron) is found within $\Delta R < 0.5$ of an AK5 jet, its four-momentum is subtracted from that of the jet; the subtraction is performed on the uncorrected jet (before the application of JECs); the energy of the resulting jet is then corrected by the appropriate JEC. A simpler jet-lepton cleaning procedure is applied for large-radius jets. Since these jets are only used for t-tagging purposes and we do not expect a top quark decaying hadronically to overlap with a prompt lepton in the present search, the CA8 jet collection is cleaned from charged leptons by requiring each jet candidate to be separated by $\Delta R > 0.8$ with respect to any of the lepton candidates; if this condition is not met by a given jet, the latter is discarded.

Jet candidates in the AK5 collection are required to have $p_T > 50$ GeV and $|\eta| < 2.4$. The CA8 jet candidates, on the other hand, have $p_T > 400$ GeV and $|y| < 2.4$, where y stands here for the jet rapidity.

Jet b-tagging

As top quarks decay almost exclusively to a W boson and a bottom quark, a final state arising from $t\bar{t}$ production is expected to be characterized, at the reconstruction level, by the presence of hadronic jets with b-flavor content. The identification of such jets, also known as jet b-tagging, thus plays a central role in discriminating $t\bar{t}$ events against other reducible backgrounds. The algorithm used in the analysis for the purpose of jet b-tagging is the Combined Secondary Vertex (CSV) algorithm [84] (see Section 3.5.3).

Small-radius jets are used to identify clusters associated to the decay of a b-quark. An AK5 jet candidate is said to be b-tagged if it passes the CSV medium operating point (CSVM); the average mis-identification rate of this b-tagging method is at the level of 1%. Data/MC corrections are applied on simulated events in order to account for discrepancies between the b-tagging efficiencies in data and MC; the implementation of this correction is described in Section 4.5.3.

Jet t-tagging

Large-radius jets are used to reconstruct the fully hadronic decay of high-p_T top quarks. The t-tagging algorithm employed in the analysis is given by the combination of two methods [87]: one is the CMS Top Tagging algorithm [34, 89], which is used to determine the subjets associated to the CA8 jet candidate, and the other is based on the N-subjettiness jet-shape variable [91, 92]. We refer to Section 3.5.4 for a detailed description of these two methods. The t-tagging working point used in this search is defined by the following requirements:

- 140 GeV $< m_{\text{jet}} <$ 250 GeV, where m_{jet} is the ungroomed mass of the CA8 jet,
- $N_{\text{subjets}} \geqslant 3$, where N_{subjets} is number of subjets reconstructed by the CMSTT algorithm,
- $m_{\text{min}} > 50$ GeV, where m_{min} is the minimum pairwise mass between the three p_T-leading subjets reconstructed by the CMSTT algorithm, and
- $\tau_{32} \equiv \tau_3/\tau_2 < 0.7$, where τ_N denotes the N-subjettiness variable defined in Equation (3.9).

A CA8 jet candidate is said to be t-tagged if it passes the selection cuts reported above. The efficiency and mis-identification rate of this t-tagging algorithm are studied in Section 4.5.4.

4.3.5 Missing transverse energy

The imbalance of the transverse energy in the event ($\not\!\!E_T$) is determined based on the momenta of all the reconstructed PF candidates [83]. The Type-1 $\not\!\!E_T$ measurement defined in Section 3.5.2, which accounts for the propagation of the jet energy scale corrections to the $\not\!\!E_T$ estimate, is used. In simulated events, the data/MC corrections for the jet energy resolution are also propagated to the $\not\!\!E_T$ measurement.

4.4 Event selection

We select events consistent with the production of a top-antitop quark pair in the semileptonic channel, where one W boson decays to a lepton (electron or muon) and a neutrino, and the other W boson decays hadronically.

The analysis selection is designed to achieve maximum efficiency for events characterized by the production of high-p_T top quarks, whose decay products can be highly collimated in the final state. The three quarks coming from the hadronically decaying top quark may be reconstructed in the detector as either three jets, two jets or even a single merged jet, depending on the top quark momentum. Similarly, in the case of the leptonic top quark decay, the lepton may be poorly isolated due to its proximity to the b-quark jet. This calls for a rather inclusive event selection, based on the requirement of one high-p_T lepton (not necessarily isolated), missing transverse energy and two or more high-p_T jets. The selection described below is applied to both data and simulated events.

4.4.1 Preselection

In the following the first part of the analysis selection, which is referred to as *preselection*, is described. This set of cuts, which are based on the kinematic properties and the multiplicities of the physics objects defined in Section 4.3, are designed to select events whose kinematics is compatible with the semileptonic decay of a $t\bar{t}$ pair.

4.4.1.1 Preselection filters and primary vertex

A set of standard reconstruction-quality criteria are applied to discard events affected by different types of detector noise or malfunctioning. Most importantly, each event is required to have at least one primary vertex candidate.

4.4.1.2 High Level Trigger

Data and MC events in the μ+jets channel are required to pass the HLT path `Mu40_eta2p1`; for real data events, this trigger belongs to the `SingleMu` data stream. Among the triggers employed in CMS during the 2012 data-taking, this is the most inclusive unprescaled single-muon trigger without any requirement on muon isolation. The HLT trigger thresholds at 40 GeV in p_T and 2.1 in absolute pseudorapidity motivate the kinematic requirements for the muon candidate considered in the analysis ($p_T > 45$ GeV and $|\eta| < 2.1$); this offline thresholds are used to select events in the plateau of the trigger efficiency.

In the electron channel, the trigger selection is given by the logical OR of two HLT paths, named `HLT_Ele30_CaloIdVT_TrkIdT_PFNoPUJet100_PFNoPUJet25_v*` and `HLT_PFJet320_v*`. The first trigger, included in the `ElectronHad` data stream, requires one non-isolated electron with $p_T > 30$ GeV and two HLT jets respectively with $p_T > 100$ GeV and $p_T > 25$ GeV; the second trigger, which belongs to the `JetHT` data stream, requires only one jet with $p_T > 320$ GeV. The electron-based trigger would be the natural candidate for the e + jets analysis, but its efficiency tends to decrease for very boosted $t\bar{t}$ decays, in which one jet and the electron are increasingly close in space and they are ultimately not resolved as two separate objects at the HLT level. This translates, for example, into a trigger efficiency lower than 90% for the highest-mass resonances ($M_X \geqslant 2$ TeV). The use of the logical OR of the "electron+2-jets" trigger with the aforementioned single-jet trigger proves to be the most effective choice to maximize the trigger efficiency for all the signal hypotheses considered in this search. As these two triggers belong to two different CMS data streams, particular care

must be taken when analyzing real data samples in order to avoid double-counting events. This is achieved by requiring the events in the `ElectronHad` data stream to pass only the "electron+2-jets" trigger and the events in the `JetHT` data stream to pass the single-jet trigger and also fail the "electron+2-jets" trigger. For simulated samples, it is sufficient to require directly the logical OR of the two triggers.

Efficiency studies in data and simulation for the HLT selection in the $\mu+$jets and $e+$jets channels are detailed in Section 4.5.1.

4.4.1.3 Lepton selection

In the muon (electron) channel, each event is required to have exactly one muon (electron) candidate, with a veto on events containing any other muon or electron candidate. A veto on additional leptons has the effect of reducing sources of background like Drell-Yan (DY) production, diboson production and $t\bar{t}$ decays in the dilepton channel.

In both channels, lepton candidates are not required to pass any cut on isolation or minimum angular separation from their nearest jet. Due to the kinematics of boosted top decays, either one of these two requirements would greatly reduce the analysis acceptance for signal events, especially for the highest-mass signal hypotheses. On the other hand, acceptance cuts (on p_T and $|\eta|$) and ID requirements alone are not enough to suppress the contribution from QCD-multijet production, which can potentially represent a major background in the single-lepton final state. In this analysis, the lepton isolation cut is replaced by a selection based on the angular distance of the lepton candidate from its closest jet and the direction of the lepton p_T: the lepton candidate is required to satisfy a cut given by the logical OR of the following two cuts

$$\Delta R_{\text{min}}(\ell, j) > 0.5 \quad \text{OR} \quad p_{T,\text{rel}}(\ell, j) > 25 \text{ GeV}\,, \tag{4.1}$$

where j denotes the AK5 jet with $p_T > 25$ GeV and the lowest ΔR-distance from the reconstructed lepton and $p_{T,\text{rel}}(\ell, j)$ stands for the lepton transverse momentum with respect to the axis of jet j. In simplest terms, if the lepton is not contained in the cone of a reconstructed jet, no cut is applied; otherwise, the lepton p_T with respect to the jet axis is required to be higher than a given threshold. In the following, this selection is referred as the *lepton 2D-cut* [117, 118]. It proves to be an effective method to reduce the QCD background without comprimising the signal efficiency. Figure 4.5 shows a comparison between the standard lepton isolation and the $p_{T,\text{rel}}(\ell, j)$ variable, for events passing all the preselection cuts except for the lepton 2D-cut; the comparison highlights the higher discrimination power of $p_{T,\text{rel}}(\ell, j)$ in distinguishing the QCD-multijet background from SM $t\bar{t}$ production and, most importantly, high-mass signal hypotheses. The two-dimensional distributions of $\Delta R_{\text{min}}(\ell, j)$ and $p_{T,\text{rel}}(\ell, j)$ can be seen in Figure 4.6, for SM $t\bar{t}$ production and two Z′ signals. These plots show that, for increasing values of the signal mass hypothesis, the average value of $\Delta R_{\text{min}}(\ell, j)$ decreases, but the majority of signal events remains in the region $p_{T,\text{rel}}(\ell, j) > 25$ GeV. The efficiency of the 2D-cut in the preselected $\mu+$jets sample equals 88% for SM $t\bar{t}$ production and it ranges between 70% and 80% for the different signal samples, whereas only 1% of the events in the QCD MC pass the selection. The lepton 2D-cut efficiency on $e+$jets events passing the preselection cuts corresponds to 95% for SM $t\bar{t}$ and it ranges between 90% and 95% for the $\text{Z}' \to t\bar{t}$ signals, while rejecting up to 60% of QCD MC events; the statistical errors on these efficiencies are below 1%. The difference in efficiency across the two channels is due to the different identification requirements applied on muons and electrons. The performance of the lepton 2D-cut has been studied in collisions data

using a control sample of dilepton events, in order to derive data/MC scale factors to correct the efficiencies in simulation; this complementary measurement is described in Section 4.5.2.

One additional requirement on the lepton is applied in the e + jets channel, in which the suppression of fakes (mis-identified jets) from QCD-multijet production is more difficult to achieve compared to the μ + jets channel. Events in the electron channel are accepted only if both the following conditions are met:

$$|\Delta\phi(e,\, \not\!\!E_T) - 1.5| < \frac{1.5}{75\ \text{GeV}} \not\!\!E_T \qquad \text{and} \qquad |\Delta\phi(j,\, \not\!\!E_T) - 1.5| < \frac{1.5}{75\ \text{GeV}} \not\!\!E_T\,, \tag{4.2}$$

where e denotes the electron candidate and j stands for the p_T-leading AK5 jet. These cuts are referred to in the following as the *triangular-cuts* [118]; their effect can be seen in Figure 4.7, which shows the distributions of $\Delta\phi(e,\, \not\!\!E_T)$ and $\Delta\phi(j,\, \not\!\!E_T)$ versus $\not\!\!E_T$, before and after the cuts in Equation (4.2). This selection leads to a further reduction of the QCD-multijet background in the e + jets channel, while leaving the signal acceptance almost unaffected. Considering the events passing the e + jets preselection (including the lepton 2D-cut), the efficiency of the triangular-cuts corresponds to 93% for SM $t\bar{t}$ production, between 94% and 96% for all the signal hypotheses and 53% for the QCD-multijet MC; the statistical uncertainties on these efficiencies are below 1%.

Figure 4.8 shows a comparison of some relevant distributions of the lepton kinematics for different MC samples (W + jets, $t\bar{t}$ and one Z′ signal) for events with one lepton candidate passing the selection described above.

4.4.1.4 Jet selection

The jet selection is based on the requirement of at least two high-p_T jets in order to cover all the possible topologies of $t\bar{t}$ semileptonic decays, from events where quark decay products are fully resolved (four or more jets) to events where the decay products of more that one quark are merged in a single jet (two or more jets). We require the p_T-leading and p_T-subleading AK5 jet candidates to have $p_T > 150$ GeV and $p_T > 50$ GeV, respectively. The jet selection is the same in the muon and electron channel. No events are discarded based on the number of b-tagged or t-tagged jets; this information is used later on in the analysis to split the final event sample in exclusive categories in order to increase the sensitivity of the search (see Section 4.4.3 and Section 4.8).

Distributions for the jet multiplicity and the kinematics of the two p_T-leading jets are shown in Figure 4.9, where the W + jets and $t\bar{t}$ backgrounds are compared to a Z′ signal. These plots are shown for the μ + jets channel; very similar distributions are obtained in the e + jets analysis.

4.4.1.5 Missing-E_T selection

The reconstructed missing E_T is required to be larger than 50 GeV. This requirement suppresses the contribution of backgrounds with no real MET in the final state, e.g. QCD-multijet and DY production. In addition, the event is required to pass a cut given by $H_T^{\text{lep}} > 150$ GeV, where H_T^{lep} corresponds to the scalar sum of the lepton p_T and the missing transverse energy. This cut has the effect of improving the overall sensitivity of the analysis for boosted $t\bar{t}$ decays and it further reduces the contamination from processes in which the lepton does not originate from a W $\to \ell\nu$ decay. The $\not\!\!E_T$ and $H_{\text{T}}^{\text{lep}}$ distributions for events passing the full preselection in the muon channel are depicted in Figure 4.10, where distributions for different simulated samples (W+jets, SM $t\bar{t}$ and one Z′ signal) are compared.

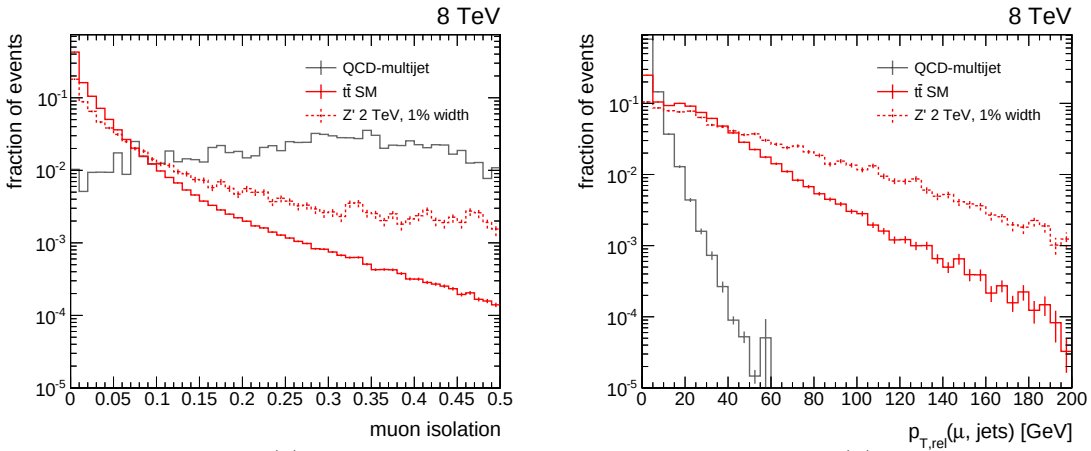


Figure 4.5: Comparison between the lepton fixed-cone isolation (a) and the $p_{T,\mathrm{rel}}(\ell, j)$ distribution (b) for simulated events passing the $\mu+$jets preselection, except for the lepton 2D-cut requirement; the $p_{T,\mathrm{rel}}(\ell, j)$ distribution includes only events with $\Delta R_{\mathrm{min}}(\ell, j) < 0.5$, which is the region potentially affected by the lepton 2D-cut. Distributions are shown for QCD-multijet production, SM $t\bar{t}$ production and a Z′ resonance with a mass of 2 TeV and 1% width. Each of the these distributions is normalized to unity.

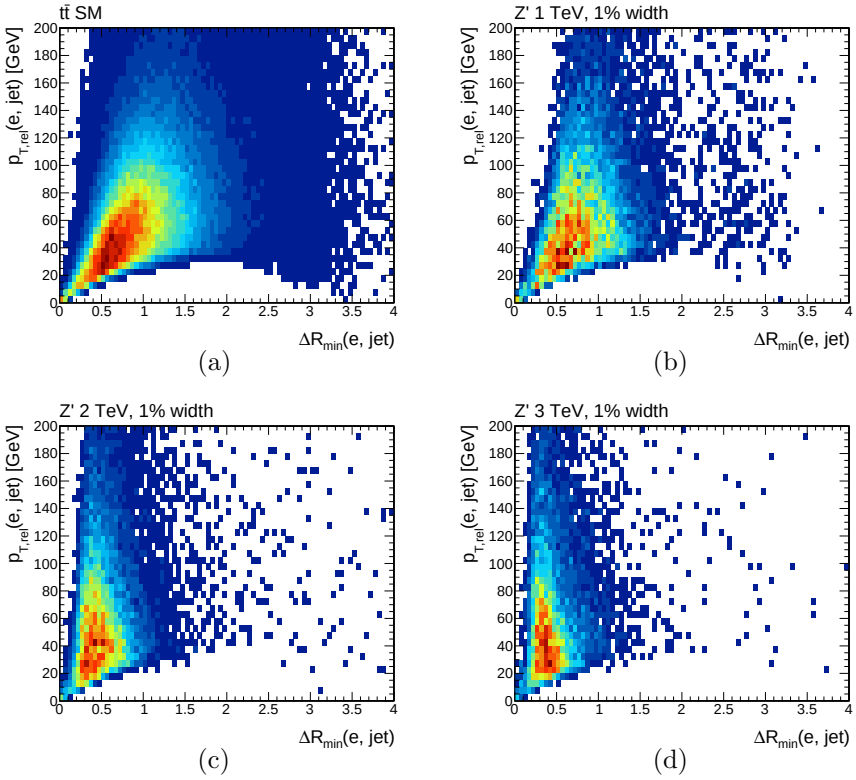


Figure 4.6: Two-dimensional distributions of $\Delta R_{\mathrm{min}}(\ell, j)$ versus $p_{T,\mathrm{rel}}(\ell, j)$ for events passing the $e+$jets preselection, except for the lepton 2D-cut requirement. Distributions are shown for SM $t\bar{t}$ production (a) and a Z′ signal with 1%-width and $M_{\mathrm{Z}'} =$ 1 TeV (b), $M_{\mathrm{Z}'} = 2$ TeV (c) and $M_{\mathrm{Z}'} = 3$ TeV (d).

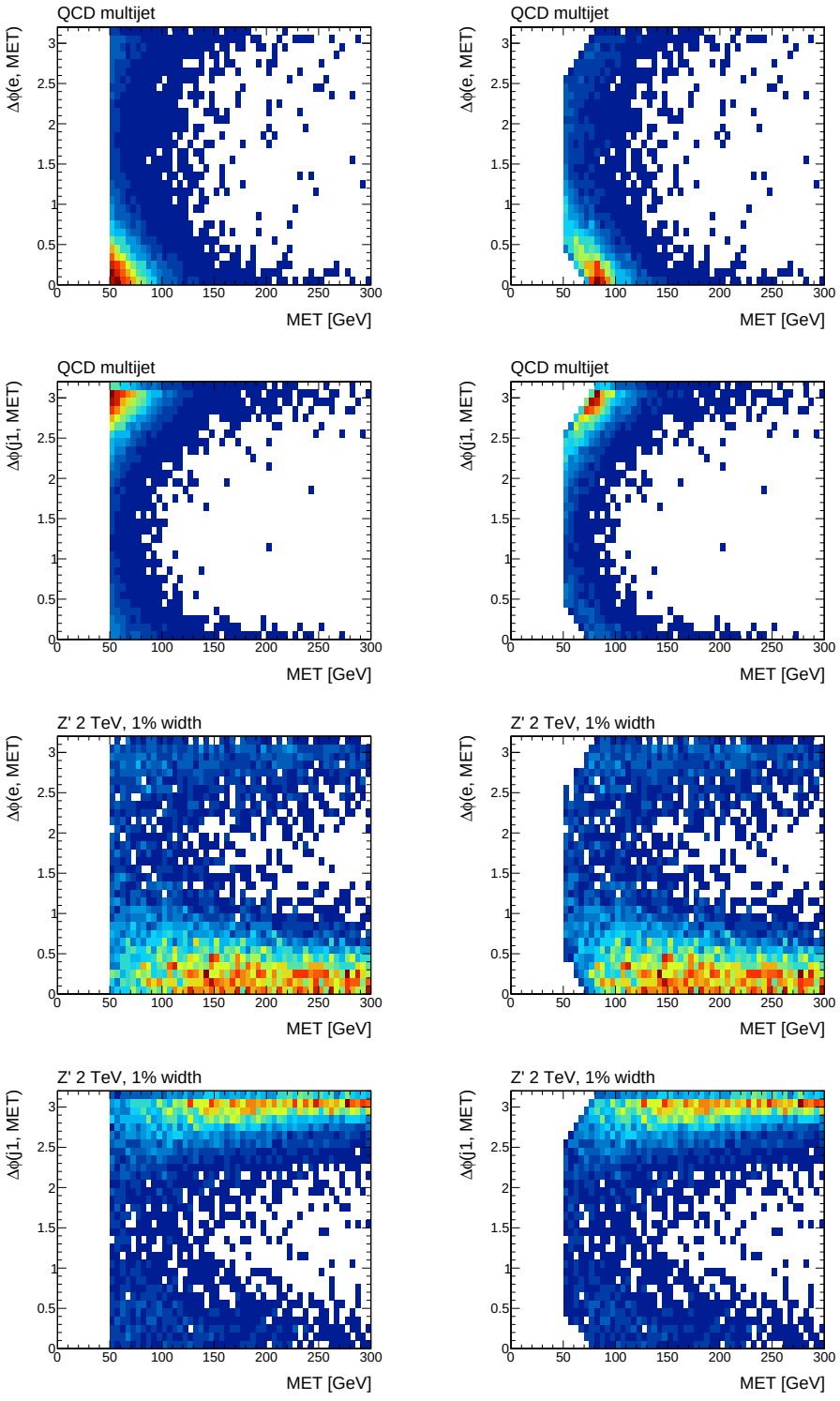


Figure 4.7: Two-dimensional distributions of $\not\!\!E_T$ versus $\Delta\phi(e, \not\!\!E_T)$ and $\Delta\phi(j1, \not\!\!E_T)$ for events passing the e + jets preselection. Plots on the left-hand side include events passing the preselection cuts, except for the triangular-cuts; plots on the right-hand side include only those events which also pass the triangular-cuts (full preselection). Distributions are shown for QCD-multijet production (upper plots) and a Z′ signal with $M_{Z'} = 2$ TeV and 1% width (lower plots).

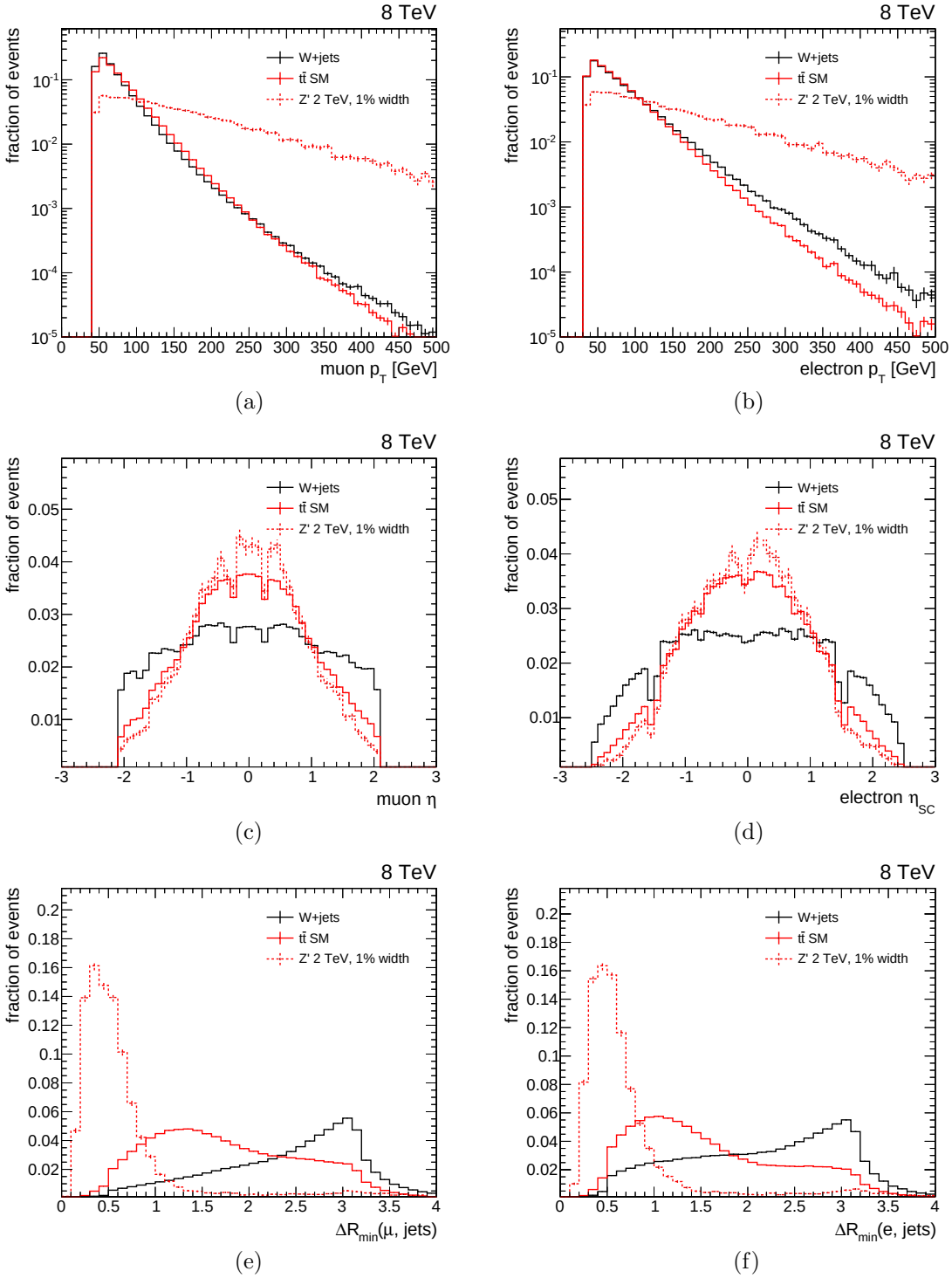


Figure 4.8: Kinematic distributions in simulated events passing the lepton selection described in Section 4.4.1.3. In each plot, histograms are shown for W+jets production, SM $t\bar{t}$ production and a Z′ resonance with a mass of 2 TeV and 1% width. Separate plots are shown for the muon and electron channels: (a) muon p_T, (b) electron p_T, (c) muon η, (d) electron supercluster η, (e) minimum ΔR between muon and jets, (f) minimum ΔR between electron and jets. Each of the these distributions is normalized to unity.

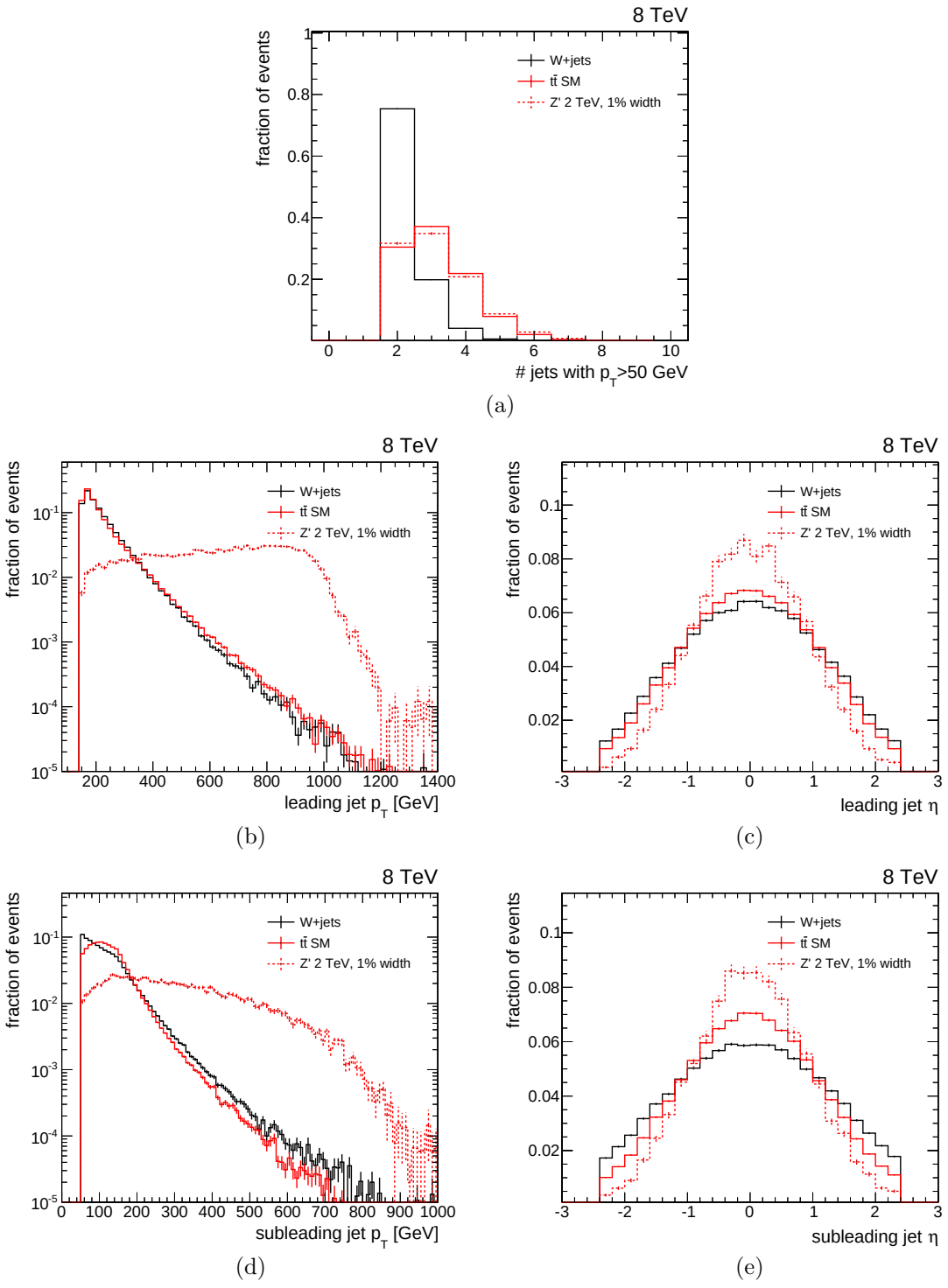
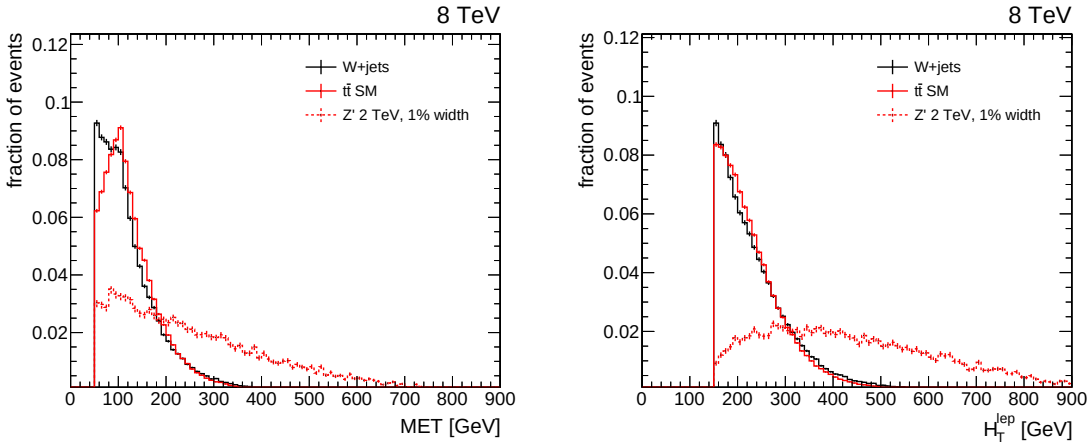


Figure 4.9: Distributions for AK5 jets in simulated events passing the lepton selection (Section 4.4.1.3) and the jet selection (Section 4.4.1.4) in the μ + jets channel: (a) number of jets with $p_T > 50$ GeV and $|\eta| < 2.4$ (b) leading jet p_T, (c) leading jet η, (d) subleading jet p_T, (e) subleading jet η. In each plot, histograms are shown for W + jets production, SM $t\bar{t}$ production and a Z′ resonance with a mass of 2 TeV and 1% width. Each of the these distributions is normalized to unity.



(a) (b)

Figure 4.10: Distributions of $\not\!\!E_T$ (a) and $H_{\mathrm{T}}^{\mathrm{lep}}$ (b) for simulated events passing the full preselection in the muon channel. In each plot, histograms are shown for W + jets production, SM $t\bar{t}$ production and a Z′ resonance with a mass of 2 TeV and 1%-width. Each of the these distributions is normalized to unity.

4.4.2 Kinematic reconstruction of the $t\bar{t}$ system

The kinematic reconstruction of the top and antitop quark four-momenta is performed for each event according to the following strategy. The charged lepton and $\not\!\!E_T$ are assigned to the leptonic top quark. AK5 jets can be assigned to either the leptonic or the hadronic top quark; in addition, if the event contains a CA8 jet passing the t-tagging working point defined in Section 4.3, the hadronic top four-momentum is reconstructed based on the kinematics of the t-tagged jet alone. For each event, all possible assignments of the reconstructed objects (lepton, $\not\!\!E_T$, jets) to the hadronic and leptonic top quarks are considered and a χ^2 discriminator is used to select exactly one hypothesis.

The missing E_T is interpreted as the transverse momentum of the neutrino in the semileptonic $t\bar{t}$ decay. The z-component of the neutrino momentum is estimated indirectly by assuming that the leptonically decaying W boson is on-shell. Under this assumption, the neutrino p_z is given by the following expression

$$p_z = \frac{\mu\, p_{z,\ell}}{p_{T,\ell}^2} \pm \sqrt{\frac{\mu^2\, p_{z,\ell}^2}{p_{T,\ell}^4} - \frac{E_\ell^2\, p_{T,\nu}^2 - \mu^2}{p_{T,\ell}^2}} \tag{4.3}$$

where p_ℓ and p_ν denote respectively the four-momenta of charged lepton and neutrino ($p_{T,\nu} \equiv \not\!\!E_T$), $M_{\mathrm{W}} = 80.39$ GeV and $\mu = M_{\mathrm{W}}^2/2 + p_{T,\ell}\, p_{T,\nu} \cos\Delta\phi_{\ell\nu}$. This quadratic equation can yield zero, one or two real solutions (N_ν^{sol}). In the absence of a real solution ($N_\nu^{\mathrm{sol}} = 0$), only the real part of the complex solution is used. If more than one solution is found ($N_\nu^{\mathrm{sol}} = 2$), both hypotheses are tested.

For events without a t-tagged jet, several hypotheses are then built based on all the possible assignments of each AK5 jet candidate to one of the top quark decays: each AK5 jet can be assigned to either the leptonic leg, the hadronic leg or none of the two (jet from QCD radiation). The number of hypotheses for each event thus corresponds to $3^{N_{\text{AK5-jet}}}$ times the number of solutions for the neutrino reconstruction, which can be one or two.

A slightly different method to reconstruct the $t\bar{t}$ hypotheses is used if a t-tagged jet is present in the event. The leptonic top reconstruction is the same as the one described above

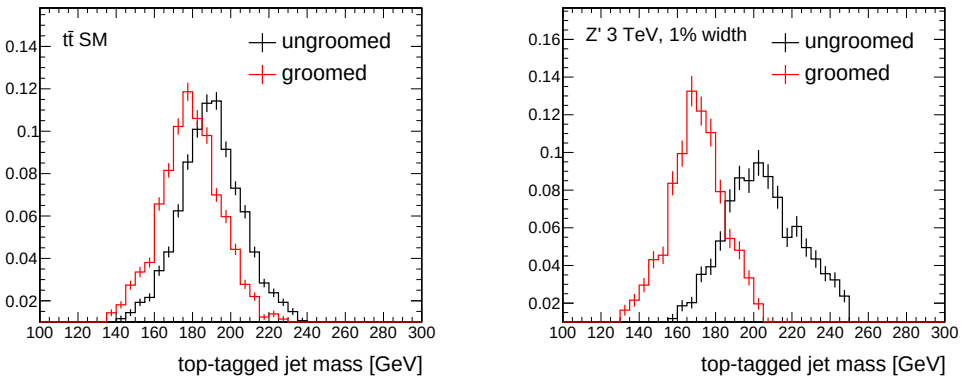


Figure 4.11: Groomed (red) and ungroomed (black) masses of a t-tagged jet for events passing the μ + jets preselection. Plots are shown for SM $t\bar{t}$ production (left) and a Z′ resonance with $M_{Z'} = 3$ TeV and $\Gamma_{Z'}/M_{Z'} = 1\%$ (right). The histograms are normalized to unity and the error bars include only the MC statistical uncertainty.

with the sole exception that only AK5 jets not overlapping with the t-tagged jet can be assigned to the leptonic leg; the no-overlap condition is given by ΔR(t-tag, AK5-jet) > 1.3. The kinematics of the hadronic top is determined based on the kinematics of the t-tagged jet. The momentum (p_x, p_y, p_z) of the hadronic top is set equal to the momentum of the CA8 jet, which is correctly calibrated due to the use of the JECs. For the mass of the hadronic top quark we use the groomed mass of the CA8 jet, which corresponds to the mass of the four-momentum sum of the subjets of the t-tagged jet. The choice of the groomed jet mass, instead of the ungroomed mass, is motivated by the fact that the groomed mass gives better performance in the reconstruction of the resonance mass, both in resolution and mean value, compared to the plain jet mass. The latter yields on average higher values than the top quark mass because of extra-radiation contained in the jet, which is removed by the grooming procedure. Summarizing, we use the jet three-momentum and the jet groomed mass as they represent the best information available to reconstruct the kinematics of the hadronic top quark decay. A comparison between the ungroomed and groomed mass of a t-tagged jet in events passing the full analysis selection is shown in Figure 4.11 for the $t\bar{t}$ MC and a Z′ signal.

Among all the possible reconstruction hypotheses, we consider only those in which at least one AK5 jet (or exactly one CA8 t-tagged jet) is assigned to the hadronic top decay and at least one AK5 jet is assigned to the leptonic top decay. The best hypothesis for each event is identified as the one that minimizes the χ^2 discriminator given by

$$\chi^2 = \left[\frac{M_{\text{top}}^{\text{lep}} - \bar{m}_{\text{top}}^{\text{lep}}}{\sigma_M^{\text{lep}}}\right]^2 + \left[\frac{M_{\text{top}}^{\text{had}} - \bar{m}_{\text{top}}^{\text{had}}}{\sigma_M^{\text{had}}}\right]^2, \tag{4.4}$$

where the four parameters $\bar{m}_{\text{top}}^{\text{lep}}$, $\bar{m}_{\text{top}}^{\text{had}}$, σ_M^{lep} and σ_M^{had} are tuned using MC events. These parameters correspond to the mean and width returned by two independent gaussian fits on the reconstructed $M_{\text{top}}^{\text{lep}}$ and $M_{\text{top}}^{\text{had}}$ distributions in MC events. The $t\bar{t}$ hypothesis used for these fits is the one satisfying a matching condition between the reconstructed objects and the decay products of the generated $t\bar{t}$ pair; this is done to remove hypotheses characterized by incorrect jet assignments or imprecise neutrino reconstruction. Specifically, we consider

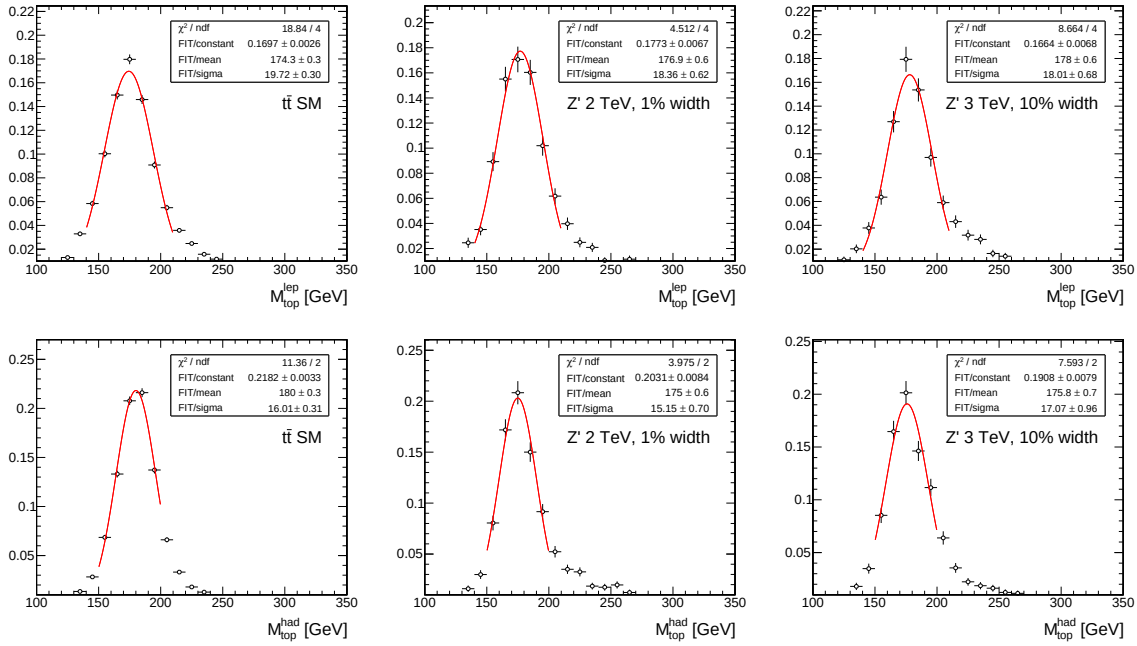


Figure 4.12: Gaussian fits of the masses of the reconstructed leptonic (top) and hadronic (bottom) top quarks for the correct $t\bar{t}$ hypothesis in simulated events. For the leptonic (hadronic) top quark mass, the fit is performed in the mass range from 140 GeV (150 GeV) to 210 GeV (200 GeV). The distributions include events passing the μ+jets preselection. Distributions are shown, from left to right, for SM $t\bar{t}$ production, a Z′ resonance with $M_{Z'} = 2$ TeV and $\Gamma/M = 1\%$ and a Z′ resonance with $M_{Z'} = 3$ TeV and $\Gamma/M = 10\%$. The templates are normalized to unity and the fit results are given in the top-right box of each plot.

the hypothesis which minimizes the ΔR sum between reconstructed objects (lepton, $\not\!\!E_T$, jets) and the corresponding generated particles (lepton, neutrino, quarks) in the MC; this hypothesis is denoted in the following as the *correct* $t\bar{t}$ hypothesis in MC events, and it is the one used to tune the parameters of the χ^2 discriminator. Figure 4.12 shows the $M_{\text{top}}^{\text{lep}}$ and $M_{\text{top}}^{\text{had}}$ distributions of the correct $t\bar{t}$ hypothesis for SM $t\bar{t}$ and two Z′ signals. The gaussian parameters of the corresponding fits are reported in the plots. The mass mean and resolution taken from the fits are found to be compatible for different signal mass hypotheses and across different signal models; these values are also in agreement with the ones obtained for SM $t\bar{t}$ production. The average mean and width measured in different MC samples are ultimately used to tune the χ^2 discriminator. The values obtained are the following: $\bar{m}_{\text{top}}^{\text{lep}} = 174$ GeV, $\bar{m}_{\text{top}}^{\text{had}} = 181$ GeV, $\sigma_M^{\text{lep}} = 18$ GeV and $\sigma_M^{\text{had}} = 15$ GeV.

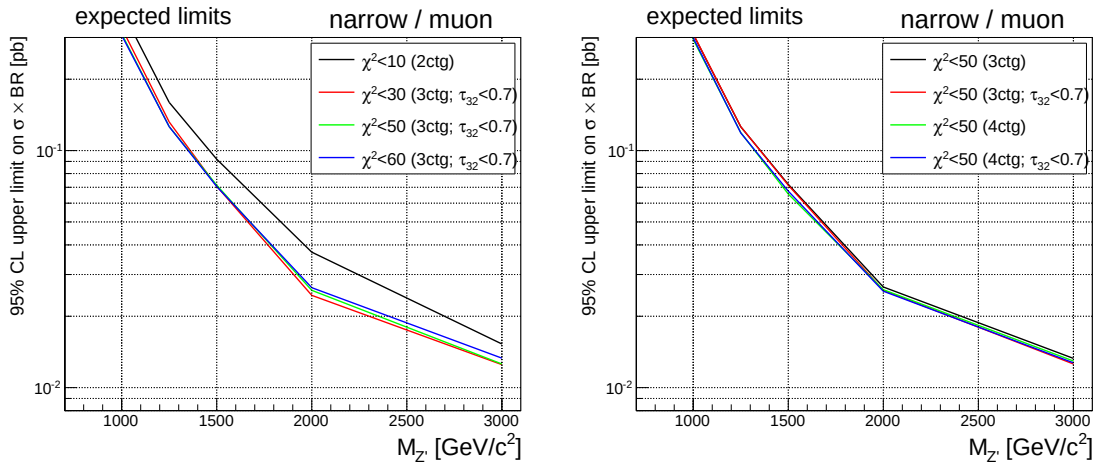


Figure 4.13: 95% CL expected limits on the production cross section times branching ratio of a narrow Z′ resonance decaying to a $t\bar{t}$ pair in the μ + jets channel for different choices of the χ^2 upper threshold and the SR categories. Legend: "ctg-2" (two categories): no b-tags, at least one b-tag; "ctg-3" (three categories): no t-tags and no b-tags, no t-tags and at least one b-tag, one t-tag; "ctg-4" (four categories): no t-tags and no b-tags, no t-tags and at least one b-tag, one t-tag and no b-tags, one t-tag and at least one b-tag. In the case of "ctg-4", the number of b-tags is given by the sum of b-tagged AK5 jets and b-tagged subjets of the t-tagged jet, if present.

4.4.3 Final event selection and categorization

After the determination of the best $t\bar{t}$ hypothesis in the event, we apply a final selection to determine the ℓ + jets signal region (SR). This selection is based on an upper cut on the χ^2 discriminator of the $t\bar{t}$ hypothesis; such a cut has the effect of reducing the contribution of non-$t\bar{t}$ backgrounds in the ℓ + jets SR and improve the sensitivity of the analysis. The optimal threshold of the χ^2 cut is estimated based on the expected exclusion limits on $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$ for a Z′ signal with 1% relative width. The same approach is used to find a categorization of the final ℓ + jets SR which maximizes the analysis sensitivity. Splitting the final SR in exclusive categories is found to be an effective method to improve the sensitivity of the analysis. As a rule, the sensitivity of a search for new physics signals can be enhanced if, instead of considering only one inclusive sample, the latter is divided in separate subsamples characterized by different background content and signal purity. The statistical combination of independent categories allows to optimally constrain the uncertainties on different backgrounds and to best assess the consistency of an excess with a BSM signal.

Figure 4.13 includes a comparison of the cross section limits obtained for different choices of the χ^2 cut and the SR categories. The expected limits for this optimization study are calculated at 95% confidence level (CL) with the same Bayesian statistical method described in Section 4.8. As shown in Figure 4.13(a), a clear improvement in the sensitivity is observed when separating events based not only on the number of b-tagged jets, but also on the number of t-tagged jets; as a matter of fact, the introduction of the event category with one t-tagged jet drives the analysis sensitivity, especially for the high-mass signal hypotheses, for which jet t-tagging becomes increasingly efficient. The same plot shows that, irrespective of the signal mass hypothesis, the expected limits are mildly dependent on the specific χ^2 thresholds, for

values of it between 30 and 50; in this range, we choose the looser selection ($\chi^2 < 50$) in order to increase the overall statistics in the ℓ + jets SR sample. An additional comparison is provided in Figure 4.13(b), where the three-categories configuration already considered in Figure 4.13(a) is compared to a more complex categorization, in which the t-tag category is divided in two subsamples based on the presence of of b-tagged AK5 jets or b-tagged subjets. Similar expected limits are found in the two cases and, thus, the simpler categorization is preferred, as the alternative would also require the introduction of an additional systematic uncertainty related to the efficiency of subjet b-tagging. Figure 4.13(b) also shows that, for either of the two categorizations, the removal of the τ_{32} cut in the t-tagging selection does not change significantly the expected limits; the τ_{32} cut is thus retained to minimize the contribution of non-$t\bar{t}$ backgrounds in events with a t-tagged jet.

Based on this optimization study, the ℓ+jets SR is defined by requiring $\chi^2_{\text{min}} < 50$, where χ^2_{min} corresponds to the χ^2 discriminator of the best $t\bar{t}$ hypothesis in the event, and the final SR categorization is given by the following three exclusive samples:

- `T1` category: events with a t-tagged jet;
- `T0B1` category: events without a t-tagged jet and with at least one b-tagged jet;
- `T0B0` category: events without a t-tagged jet and without b-tagged jets.

The latter categorization is applied separately in each lepton channel (μ + jets and e + jets), leading to a total of six event categories.

In the e + jets channel, we apply one additional cut given by $p_T^{\text{t-lep}} > 140$ GeV, where $p_T^{\text{t-lep}}$ denotes the transverse momentum of the reconstructed leptonic top quark. This cut has been designed to ultimately reduce the QCD-multijet contribution in the e+jets sample down to a negligible amount.

Signal efficiencies

Table 4.2 shows the selection efficiencies in each SR category for different mass hypotheses of the three signal models. The efficiency reported corresponds to the number of events in the MC sample passing the full analysis selection, divided by the total number of generated events (which include all $t\bar{t}$ decay modes).

Table 4.2: Overall signal efficiency of the full analysis selection for mass hypotheses of $M_X = 1$ TeV, $M_X = 2$ TeV and $M_X = 3$ TeV. Each efficiency corresponds to the percentage of MC events which pass all the cuts of the ℓ+jets selection; generated events include all the $t\bar{t}$ decay modes. Values are shown separately for each event category of the ℓ + jets SR.

Signal efficiency [%]	μ + jets			e + jets		
	T1	T0B1	T0B0	T1	T0B1	T0B0
Z′ 1% width ($M = 1$ TeV)	0.5	2.3	0.4	0.5	2.5	0.5
Z′ 1% width ($M = 2$ TeV)	2.4	2.5	1.1	2.6	2.7	1.1
Z′ 1% width ($M = 3$ TeV)	2.1	2.4	1.6	1.9	2.4	1.4
Z′ 10% width ($M = 1$ TeV)	0.5	2.1	0.4	0.5	2.3	0.4
Z′ 10% width ($M = 2$ TeV)	2.0	2.3	0.9	2.0	2.5	0.8
Z′ 10% width ($M = 3$ TeV)	1.4	1.9	0.9	1.5	2.2	0.8
g_{KK} ($M = 1$ TeV)	0.4	1.9	0.4	0.3	1.9	0.4
g_{KK} ($M = 2$ TeV)	1.3	1.9	0.7	1.4	2.0	0.6
g_{KK} ($M = 3$ TeV)	0.9	1.5	0.5	0.9	1.6	0.5

4.5 Efficiency studies in data and simulation

4.5.1 Trigger efficiency

4.5.1.1 Muon channel

The HLT path `HLT_Mu40_eta2p1_v*` is used to trigger events in the μ+jets sample. Since the offline selection requires the presence of a muon candidate with $p_T > 45$ GeV and $|\eta| < 2.1$, the phase space considered in the analysis is expected to be in the plateau of this trigger selection. Figure 4.14 shows the trigger efficiency for different MC samples in events passing the preselection cuts described in Section 4.4.1. The overall trigger efficiency for SM $t\bar{t}$ production is expected to be around 92%. For the signal MC samples, the efficiency ranges from 92% for lower-mass resonances ($M_{Z'} \lesssim 1$ TeV) to 88% for higher-mass signal hypotheses; this is due to a slight decrease of the trigger efficiency for poorly isolated muons and for muons characterized by very high p_T values, for which the impact of radiative processes becomes sizeable. The efficiencies of the `HLT_Mu40_eta2p1_v*` trigger has been measured in data with the tag-and-probe method using a sample dominated by $Z \to \mu^+\mu^-$ production [75]. The corresponding data/MC correction factors, measured in different bins of the muon p_T and η, are applied on MC events in order to correct the muon trigger efficiency in the simulation.

4.5.1.2 Electron channel

The trigger strategy employed in the e+jets analysis is based on the use of two independent triggers, an electron-based trigger and a jet-based trigger.

First, we consider a trigger seeded by one electron candidate and two jet candidates; specifically, we consider the `HLT_Ele30_CaloIdVT_TrkIdT_PFNoPUJet100_PFNoPUJet25_v*` HLT path (referred to as `HLT_Ele30[...]` in the following, for brevity), which requires the presence of one electron candidate with p_T larger than 30 GeV and two jets with $p_T > 100$ GeV and $p_T > 25$ GeV, respectively; this trigger does not include any requirement on the electron isolation. The jet requirements at trigger level were necessary to maintain this HLT path unprescaled during the 2012 data-taking campaign without raising the electron p_T-threshold above 30 GeV. This is the same trigger used in a previous version of this analysis published by the CMS Collaboration [119,120]. The efficiency of the `HLT_Ele30[...]` trigger for different Z′ MC signals is given in Figure 4.15. The plot shows that the "electron+2-jets" trigger alone is increasingly inefficient for higher values of resonance's mass. This is due to the fact that, as the p_T of the top quark candidate increases, the electron and the b-jet coming from the top decay are closer and closer to each other in space and increasingly less likely to be reconstructed as two separate objects at trigger level (no specific jet-lepton cleaning is applied at HLT); as a consequence, these events are very similar to dijet events in the online reconstruction and they are rejected by the "electron+2-jets" trigger. This effect can be seen, for example, by looking at the the efficiency of the "electron+2-jets" trigger as a function of the angular distance between the electron and its nearest jet, which is displayed in Figure 4.15(b); this efficiency drops when electron and jet are too close to each other and thus not reconstructed as separate objects at trigger level.

In order to improve the trigger efficiency for boosted $t\bar{t}$ events in the e + jets analysis, we choose to use the logical OR of the "electron+2-jets" trigger and an independent jet-based trigger, corresponding to the `HLT_PFJet320_v*` HLT path, an unprescaled single-jet trigger with an online jet p_T-threshold of 320 GeV. As can be seen in Figure 4.15(a), the inclusion of the single-jet trigger allows to improve the overall trigger efficiency up to 15% for the highest-mass signal MC samples, whereas it has a limited impact for lower-mass

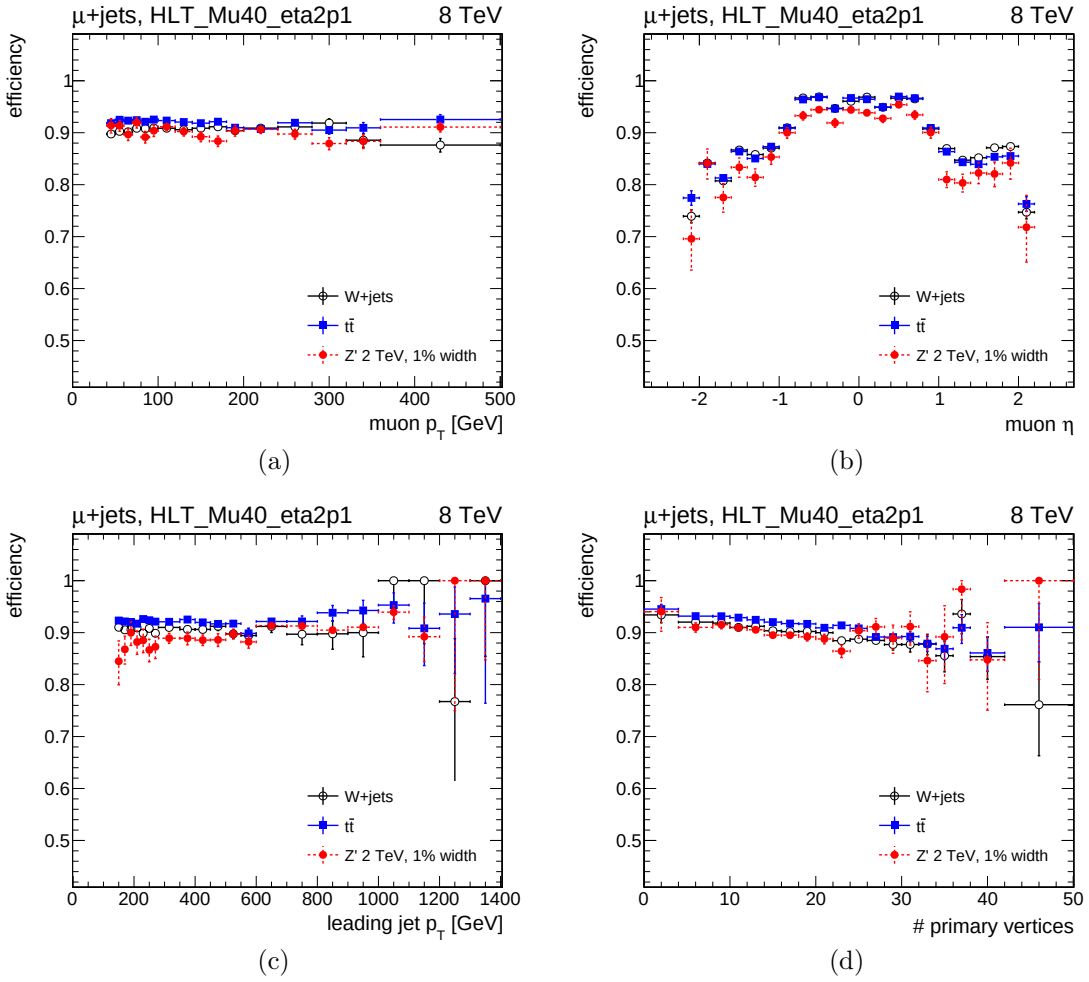
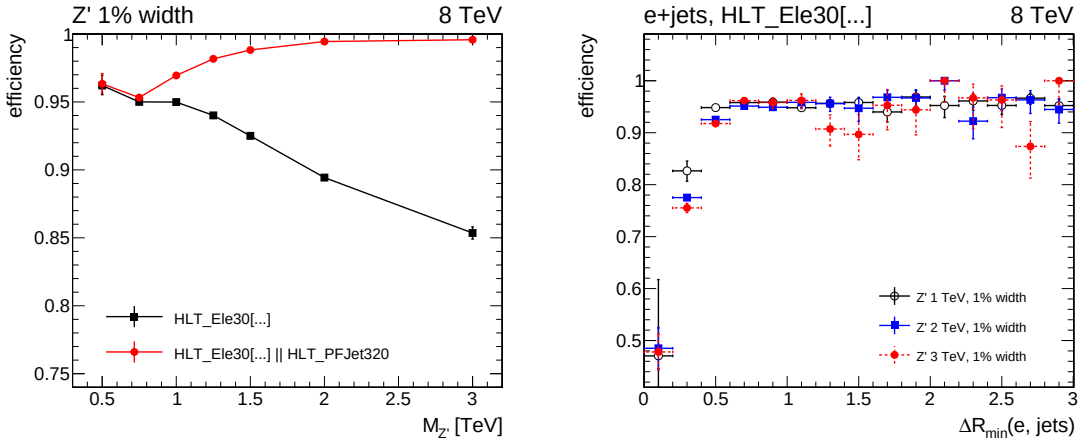


Figure 4.14: Efficiency of the single-muon trigger path `HLT_Mu40_eta2p1_v*`, as a function of different kinematic quantities, for events passing the μ + jets preselection: (a) muon p_T, (b) muon η, (c) AK5 leading jet p_T and (d) number of reconstructed primary vertices. In each plot, efficiencies are shown for W + jets, $t\bar{t}$ and a Z′ signal with $M_{Z'} = 2$ TeV and $\Gamma_{Z'}/M_{Z'} = 1\%$.

resonances ($M_X \lesssim 1$ TeV), due to the relatively high jet p_T-threshold. The overall efficiency of the combined trigger is expected to stay above 95% for all the signal mass hypotheses considered in the analysis. Based on simulation, the total yield of the SM backgrounds in the ℓ + jets sample is expected to increase only up to 2% with this new trigger strategy, compared to using the "electron+2-jets" trigger alone.

Trigger efficiency measurement in $e\mu$ events

The efficiency of the combined trigger used in the e + jets analysis is measured in data using a $e\mu$ sample dominated by $t\bar{t}$ production. This dilepton control sample is triggered by the `HLT_Mu40_eta2p1` trigger and the event selection is defined by taking the μ + jets preselection in Section 4.4 without the electron veto and adding the requirement of exactly one electron passing the kinematic cuts listed in Section 4.3; in addition, an opposite-charge condition is applied to the resulting $e\mu$ pair. The sample is dominated by $t\bar{t}$ production



(a) (b)

Figure 4.15: (a) Overall efficiencies of the `HLT_Ele30[...]` trigger (black) and the logical OR of the `HLT_Ele30[...]` and `HLT_PFJet320_v*` triggers (red) for the narrow-width Z′ signals, as a function of the Z′ mass hypothesis, for events passing the $e+$jets preselection. (b) Efficiency of the `HLT_Ele30[...]` trigger as a function of the ΔR-distance between the electron and its nearest AK5 jet for events passing the $e+$ jets preselection.

($\sim$ 91%) with small contributions from single top production and electroweak processes ($\sim$ 9%). A data/MC comparison for some relevant kinematic quantities in this control sample is displayed in Figure 4.16.

The overall efficiency of the combined trigger is $90.0 \pm 0.7\%$ in data and $95.1 \pm 0.2\%$ in the $t\bar{t}$ MC; the corresponding values for the "electron+2-jets" trigger alone are $88.5 \pm 0.7\%$ (data) and $94.1 \pm 0.2\%$ ($t\bar{t}$ MC). The error on each of these efficiencies includes only the MC statistical uncertainty. As expected, the use of the logical OR with the `HLT_PFJet320_v*` trigger has only a mild effect on the trigger efficiency of data and MC backgrounds; in particular, the event yield for the sum of all SM backgrounds in the $e\mu$ sample increases by less than 2% using the combined trigger instead of the "electron+2-jets" trigger alone. As can be seen in Figure 4.17, the observed efficiency of the combined trigger is flat, both in data and simulation, as a function for all the kinematic variables considered, except for the leading jet p_T and the PF-based electron isolation: for the former, the dependence is justified by the inclusion of the single-jet trigger; for the latter, the drop in efficiency is driven by the "electron+2-jets" trigger, which is less efficient when the electron is poorly isolated. The data/MC ratio of the efficiency, measured using the $t\bar{t}$ MC, remains flat for all the kinematic variables considered, except for the leading jet p_T, as shown in Figure 4.18.

The kinematics of the SM $t\bar{t}$ is very different from that of high-mass signals, as indicated by the different gain of these samples in terms of overall efficiency after the inclusion of the `HLT_PFJet320_v*` trigger. Therefore, in order to make a fair comparison of the trigger response of these MC samples and study possible differences, we measure their trigger efficiency in a narrower region of phase space, where the samples are kinematically more similar. To achieve this, we increase the cut on the leading jet p_T from 150 GeV to 300 GeV. Figure 4.19 shows the ratio between the trigger efficiency of SM $t\bar{t}$ and a 2 TeV narrow-width Z′ signal, for events in which the leading jet has $p_T > 300$ GeV, as a function of electron PF-isolation and the minimum ΔR between electron and jets. For all the variables considered, the fit of these ratios with a constant function is in agreement with unity within 1%. This justifies the

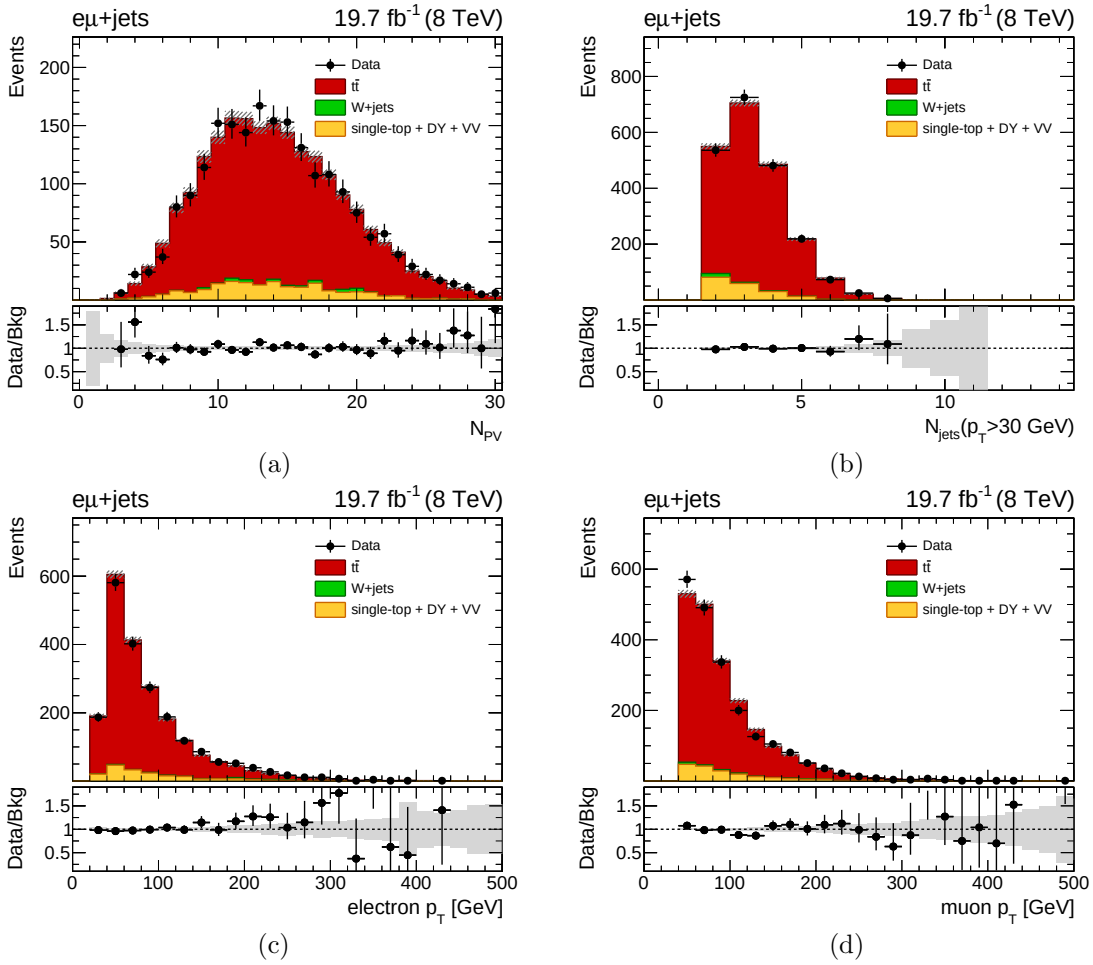


Figure 4.16: Data/MC comparisons of different kinematic variables in the $e\mu$ control sample: (a) number of primary vertices, (b) number of AK5 jets with $p_T > 30$ GeV, (c) electron p_T, (d) muon p_T. The sum of the MC backgrounds is normalized to the data. The gray band corresponds to the statistical uncertainty of the simulation. The bottom plot of each figure shows the data-over-background ratio.

use of the same correction factor for the trigger efficiency in background and signal MCs.

We measure a data/MC scale factor for the efficiency of the logical OR of the two triggers using the $t\bar{t}$ MC and we parameterize it as a function of the leading jet p_T. The fit function used is given by

$$f(p_T) = D + \frac{C}{2}\left[1 + \mathrm{erf}\left(\frac{p_T - A}{\sqrt{2}B}\right)\right] \tag{4.5}$$

and it is the sum of a constant and an error function: the former reproduces the same scale factor for the "electron+2-jets" trigger alone, while the latter models the contribution of the single-jet trigger. The results are shown in Figure 4.20. This scale factor is used to correct the MC samples in the e + jets analysis. The $\pm 1\sigma$ band on the fit results is obtained by propagating the full fit covariance matrix and it ranges from 0.7% ($p_T^{j1} < 350$ GeV) to 1.5% ($p_T^{j1} < 400$ GeV) as a function of the leading jet p_T. Since the `HLT_PFJet320_v*` is known to reach full efficiency in the high-p_T region (above 400 GeV), we assign a flat 1% systematic

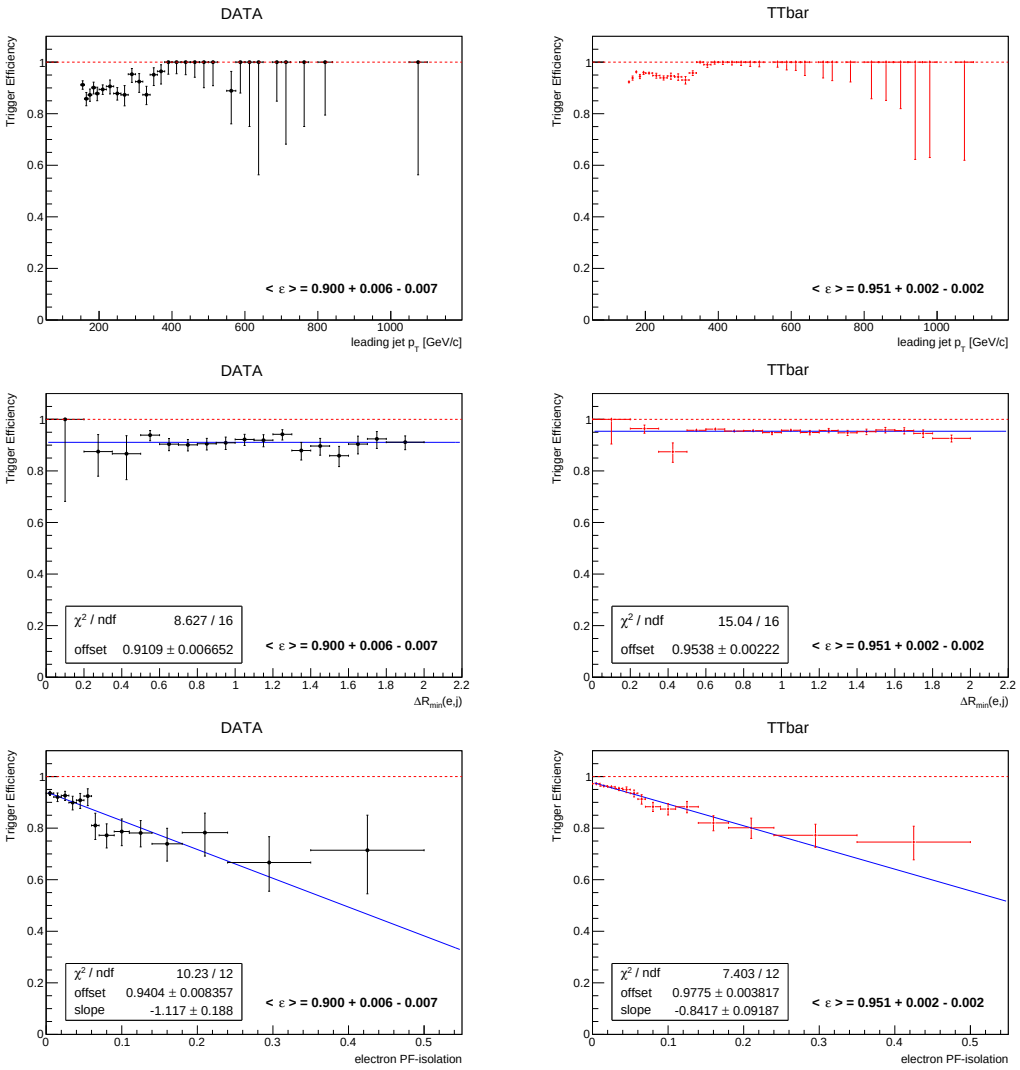


Figure 4.17: Efficiency of the logical OR of the "electron+2-jets" and `HLT_PFJet320_v*` triggers for data (left) and $t\bar{t}$ MC (right) shown, from top to bottom, as a function of leading AK5 jet p_T, minimum ΔR distance between electron and AK5 jets and electron PF-isolation. The efficiencies with respect to $\Delta R_{\mathrm{min}}(e, j)$ (electron PF-isolation) are fitted with a constant (linear) function for both data and MC.

uncertainty on the trigger scale factor. As a closure test, we measure the data/MC efficiency ratios (data over $t\bar{t}$ MC) for the combined trigger after applying the aforementioned SF to the simulation. For all the kinematic variables considered, these ratios are flat and in good agreement with unity, as shown in Figure 4.21.

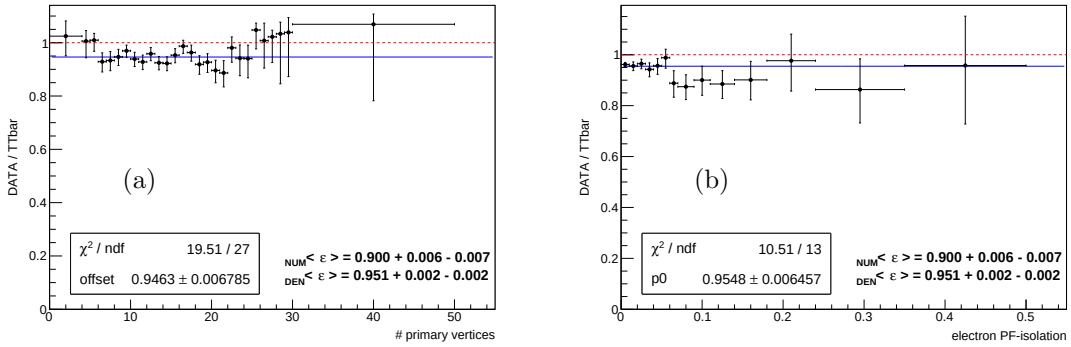


Figure 4.18: Data/MC ratio of the efficiency of the logical OR of the "electron+2-jets" and `HLT_PFJet320_v*` trigger, plotted as a function of (a) the number of primary vertices and (b) the electron PF-isolation. Efficiency fits with a constant function are also shown (blue solid lines).

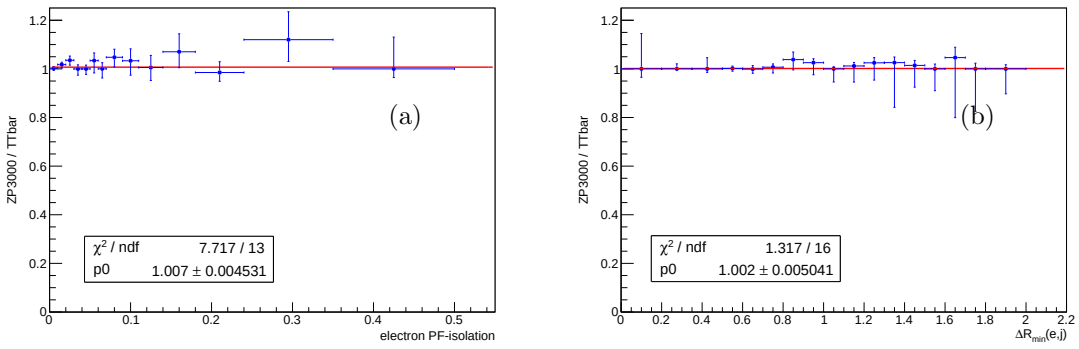


Figure 4.19: Signal/background efficiency ratios ($Z' \to t\bar{t}$ signal, with $M_{Z'} = 3$ TeV, over the $t\bar{t}$ MC) for the logical OR of the "electron+2-jets" and `HLT_PFJet320_v*` triggers, only for events with $p_T^{j1} > 300$ GeV. Efficiency ratios are plotted as a function of (a) the electron PF-isolation and (b) the minimum ΔR distance between electron and AK5 jets. In each plot, the ratio is fitted with a constant function (red solid line).

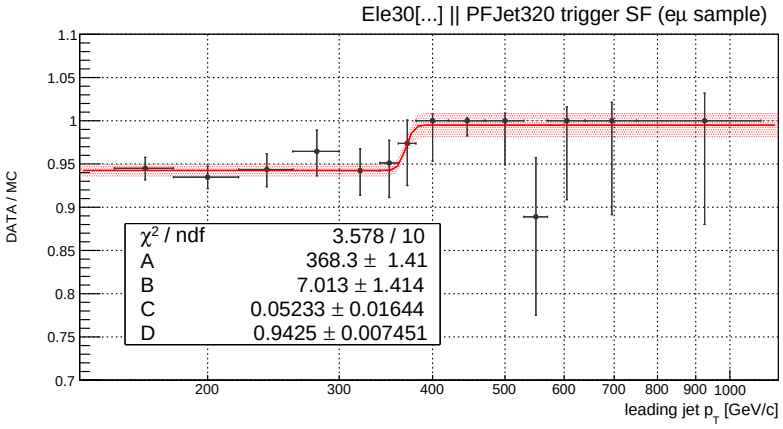


Figure 4.20: Data/($t\bar{t}$ MC) efficiency ratio for the logical OR of the "electron+2-jets" and `HLT_PFJet320_v*` triggers plotted as a function of the leading AK5 jet p_T. The red shaded area corresponds to the $\pm 1\sigma$ statistical error band obtained by propagating the full correlation matrix of the fit result (red solid line).

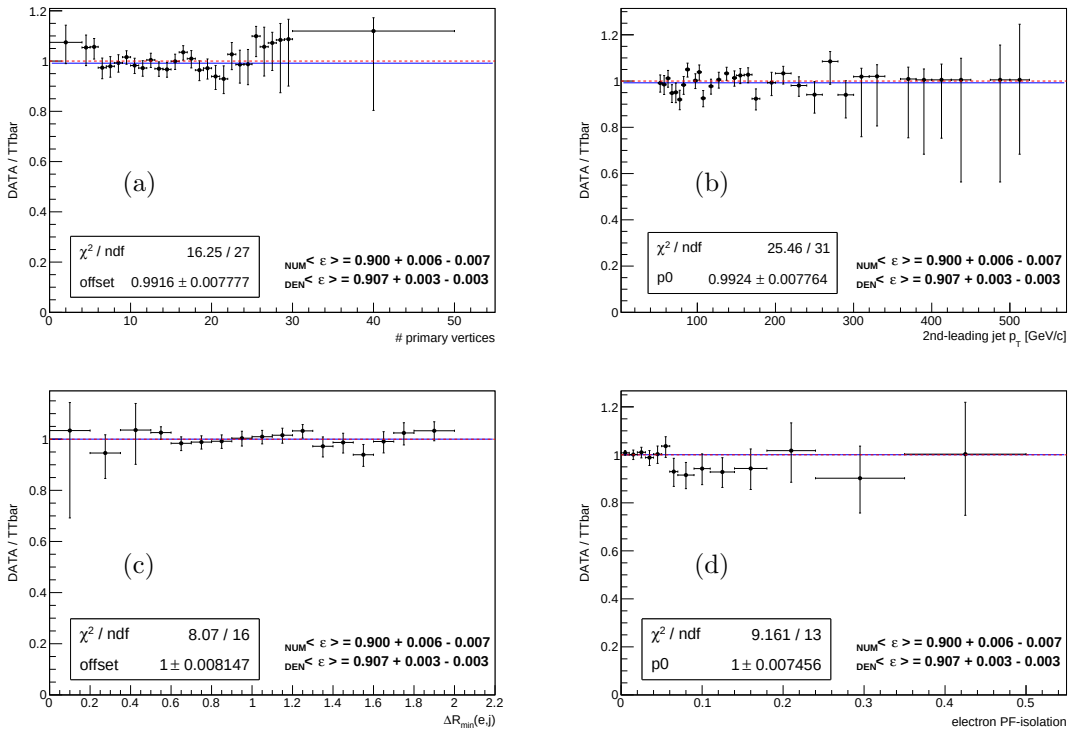


Figure 4.21: Data/($t\bar{t}$ MC) efficiency ratio for the logical OR of the "electron+2-jets" and `HLT_PFJet320_v*` triggers, after applying the p_T-dependent trigger scale factor to the simulation. Ratios are plotted as a function of (a) number of primary vertices, (b) subleading AK5 jet p_T, (c) minimum ΔR distance between electron and AK5 jets and (d) electron PF-isolation. In each plot, the ratio is fitted with a constant function (blue solid line).

4.5.2 Lepton 2D-cut efficiency

An important aspect of the ℓ+jets selection is the use of the lepton 2D-cut, defined in Equation (4.1). As discussed in Section 4.4, this cut effectively substitutes the more conventional requirement of lepton isolation, which proves to be ineffective for highly boosted semileptonic top quark decays. Applying the lepton 2D-cut, on the other hand, allows to maintain high signal efficiency and reduce the majority of the QCD-multijet background in the ℓ + jets sample. The lepton 2D-cut has been already used in previous searches for $t\bar{t}$ resonances in the ℓ + jets channel carried out by the CMS Collaboration [117, 118], but it has not been utilized frequently outside of this context.

We measure directly in data the efficiency of the lepton 2D-cut using a tag-and-probe technique [75,77] in two control regions (CR) dominated by $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events, respectively. The tag-and-probe method used to determine the lepton 2D-cut efficiencies is implemented as follows. Selected events are required to have one lepton passing tight identification requirements, i.e. the *tag* lepton, and a second lepton, referred to as *probe* lepton, which is used to measure the efficiency of interest. The invariant mass distribution of the dilepton system, $M_{\ell^+\ell^-}$, is determined separately for probe leptons passing and failing the cut under study, which, in this instance, is the lepton 2D-cut. Each of these two mass spectra is fitted to determine the contribution of signal leptons from the Z decay, which is expected to have a resonant shape, and the combinatorial background from leptons not originating from a Z boson. As a result, the cut efficiency for the probe lepton is given by

$$\varepsilon_{\mathrm{TP}} = \frac{N^{\mathrm{sig}}_{\mathrm{pass}}}{N^{\mathrm{sig}}_{\mathrm{pass}} + N^{\mathrm{sig}}_{\mathrm{fail}}} \tag{4.6}$$

where $N^{\mathrm{sig}}_{\mathrm{pass}}$ ($N^{\mathrm{sig}}_{\mathrm{fail}}$) corresponds to the total yield of the fit signal component for probes passing (failing) the selection under study. The same procedure is applied to both data and simulated events, and the resulting efficiencies are used to determine a data/MC scale factor to correct the simulation.

The dimuon and dielectron control regions are defined as follows. Events in the $\mu\mu$ CR are triggered by the isolated single-muon trigger `HLT_IsoMu24_eta2p1_v*` and they are required to have two opposite-sign muons passing $p_T > 45$ GeV, $|\eta| < 2.1$ and the muon identification cuts listed in Section 4.3. Similarly, the isolated single-electron trigger `HLT_Ele27_WP80_v*` is employed for the *ee* events, which are required to have two opposite-sign electrons with $p_T > 35$ GeV and $|\eta| < 2.5$ and passing the MVA-based identification described in Section 4.3. Events in each of the two CRs are required to have a *tag* lepton, defined as the highest-p_T lepton candidate passing $p_T > 50$ GeV, a minimum ΔR distance between tag and jets ($\Delta R > 0.5$) and a tight cut on the standard fixed-cone isolation; the latter requirement ensures that the tag selection is tighter than the cuts applied at trigger level. The higher p_T cut on the tag lepton is chosen to match the $\not{E}_T$ threshold used in the ℓ+jets analysis. These additional cuts ensure that the tag selection has a very low mis-identification rate. The other lepton is identified as the probe lepton; the kinematic cuts and identification requirements applied on the probe lepton in the dilepton CR are thus the same ones used for the lepton candidate in the ℓ + jets analysis. The control region selection includes the same jet requirements used in the ℓ+jets analysis, namely $p_T > 150$ GeV on the leading jet and $p_T > 50$ GeV on the subleading jet; these cuts are added to make the $\ell\ell$ CRs as close as possible to the ℓ + jets selection. For the same reason, we also require $H_{\mathrm{T}}^{\mathrm{lep}} \equiv p_T^{\ell 1} + p_T^{\ell 2} > 150$ GeV in the two CRs. Figures 4.22–4.23 include data/MC comparisons for some relevant kinematic quantities in the two dilepton samples defined above. For all the distributions considered,

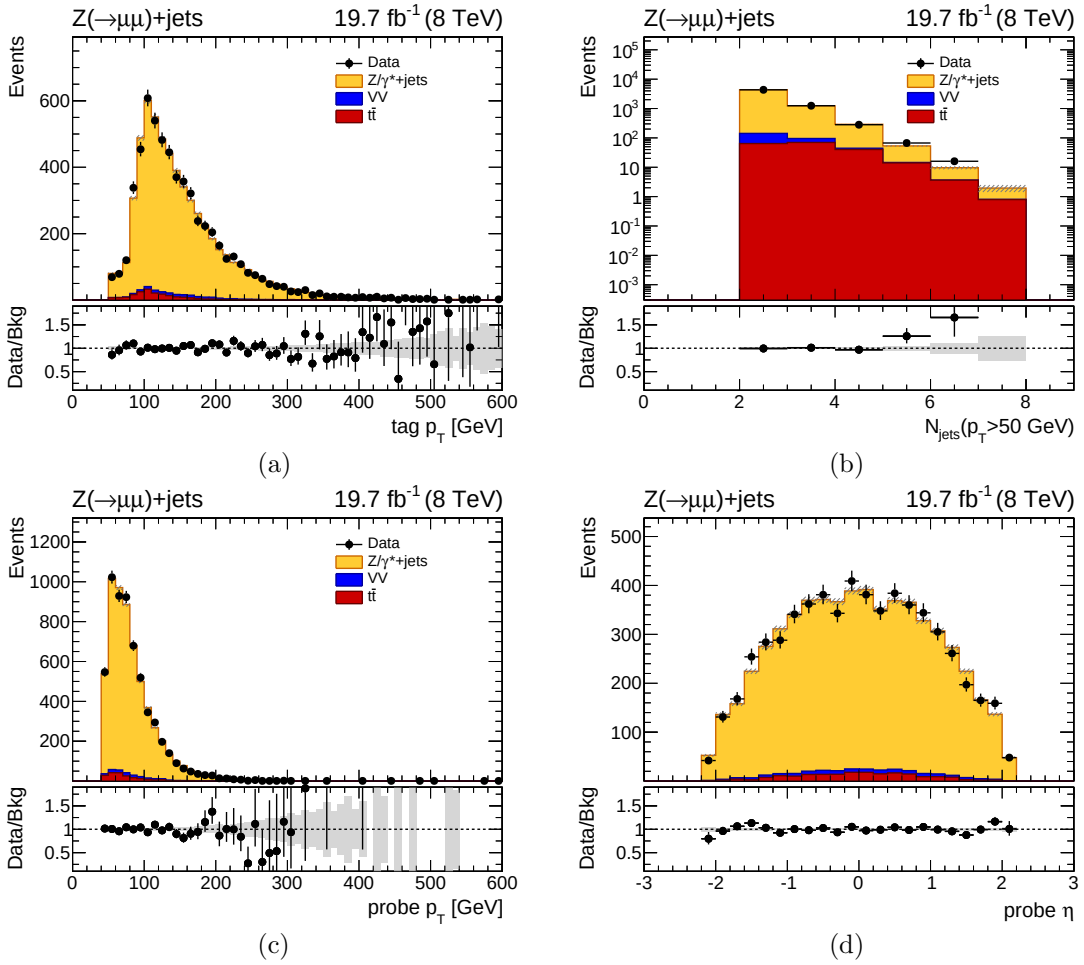


Figure 4.22: Data/MC comparisons in the dimuon CR: (a) tag muon p_T, (b) number of AK5 jets with $p_T > 50$ GeV, (c) probe muon p_T, (d) probe muon η. The total background is normalized to the data. The gray error band includes only the statistical uncertainty of the simulation. The bottom plot of each figure shows the data-over-background ratio.

the simulation provides a good description of the data in both the $\mu\mu$ and ee control regions. Both samples are dominated by $\mathrm{Z}(\rightarrow \ell^+\ell^-) + \text{jets}$ production, with a small contribution from $t\bar{t}$ and diboson events. In the following, the $\mathrm{Z} + \text{jets}$ MC is used to determine the tag-and-probe efficiencies for simulated events.

For the tag-and-probe fits of the dilepton mass distributions, we consider a mass interval from 70 GeV and 130 GeV in the $\mu\mu$ channel and from 60 GeV and 120 GeV in the ee channel. For both channels, the probability distribution function (`pdf`) for the signal component corresponds to the convolution of a Breit-Wigner distribution with a Crystal-Ball function and a falling exponential is taken as the background `pdf`. The Crystal-Ball function in the signal `pdf` is used to model the radiative tail of the Z mass peak.

Since the lepton 2D-cut effectively corresponds to a cut on $p_{T,\mathrm{rel}}$ only for events with $\Delta R_{\mathrm{min}}(\ell, \text{jets}) < 0.5$ and it has no effect outside of this region, this low-ΔR subset is the the most sensitive one to study a possible discrepancy between data and simulation due to this cut. Based on the above, we measure the lepton 2D-cut efficiency in this region as a function of $\Delta R_{\mathrm{min}}(\ell, \text{jets})$. Figure 4.24 shows the efficiencies obtained with the tag-and-probe

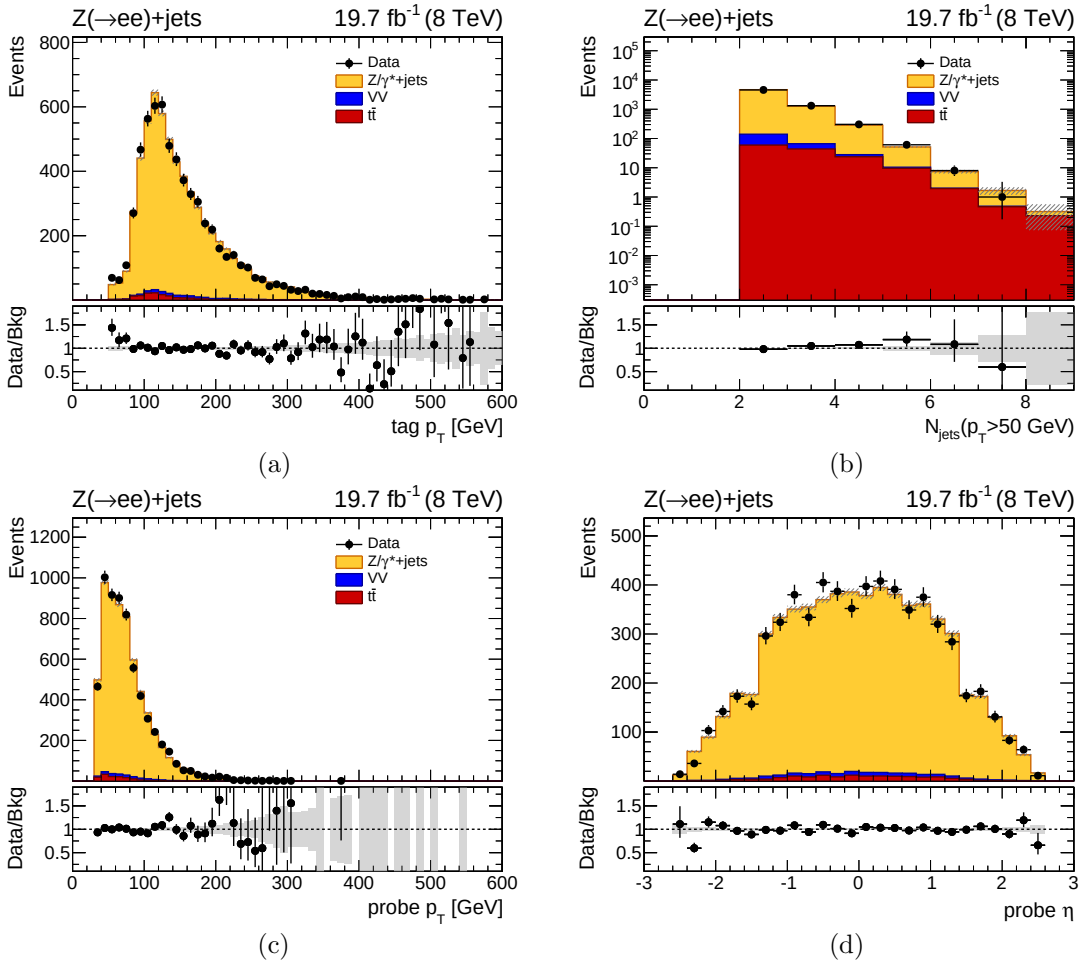


Figure 4.23: Data/MC comparisons in the dielectron CR: (a) tag electron p_T, (b) number of AK5 jets with $p_T > 50$ GeV, (c) probe electron p_T, (d) probe electron η. The total background is normalized to the data. The gray error band includes only the statistical uncertainty of the simulation. The bottom plot of each figure shows the data-over-background ratio.

method in the two dilepton CRs; the corresponding data/MC efficiency ratios can be seen in Figure 4.24. The mass fits used to derive these efficiencies can be found in Figures 4.26–4.29. The lepton 2D-cut efficiencies in data and MC are found to be in agreement with each other within their uncertainties. The ratio of these efficiencies is fitted in order to determine a data/MC SF for the lepton 2D-cut efficiency as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$. A linear function is chosen to perform the SF fit and, since the lepton 2D-cut efficiency equals unity by definition for $\Delta R_{\text{min}} > 0.5$, we also require $\text{SF}(\Delta R_{\text{min}} = 0.5) = 1$. Based on this, the fit function finally corresponds to

$$f(\Delta R_{\text{min}}) = 1 + \alpha(2\,\Delta R_{\text{min}} - 1) \tag{4.7}$$

which is linear in ΔR_{min} and only depends on the α parameter. The fit is performed separately for the muon and electron channels. The fit results, shown in Figure 4.25, are $\alpha_\mu = 0.11 \pm 0.25$ and $\alpha_e = -0.12 \pm 0.31$, and they are in agreement across the two channels. These fit functions are used in the μ + jets and e + jets analyses, respectively, to correct the lepton 2D-cut efficiency in simulated events.

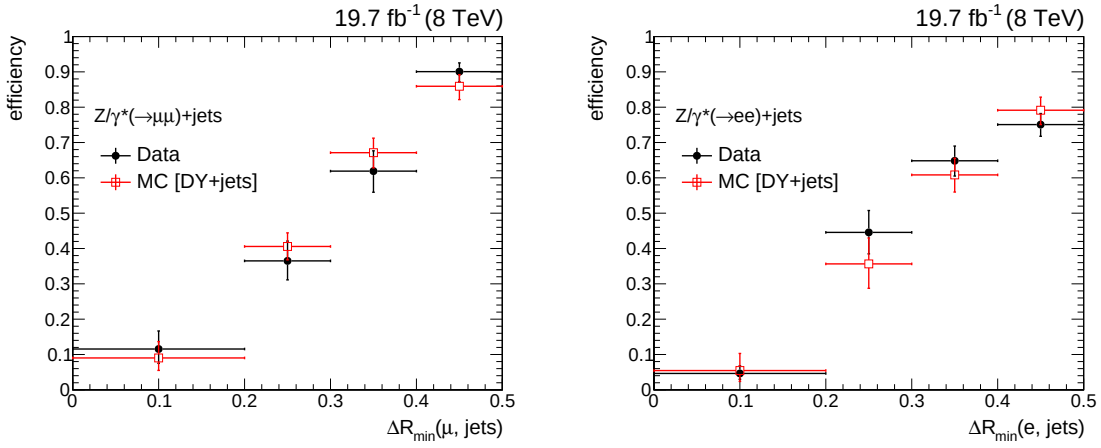


Figure 4.24: Lepton 2D-cut efficiency, as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$, for data and simulated events in the $\ell^+\ell^-$ CRs: (left) dimuon sample, (right) dielectron sample.

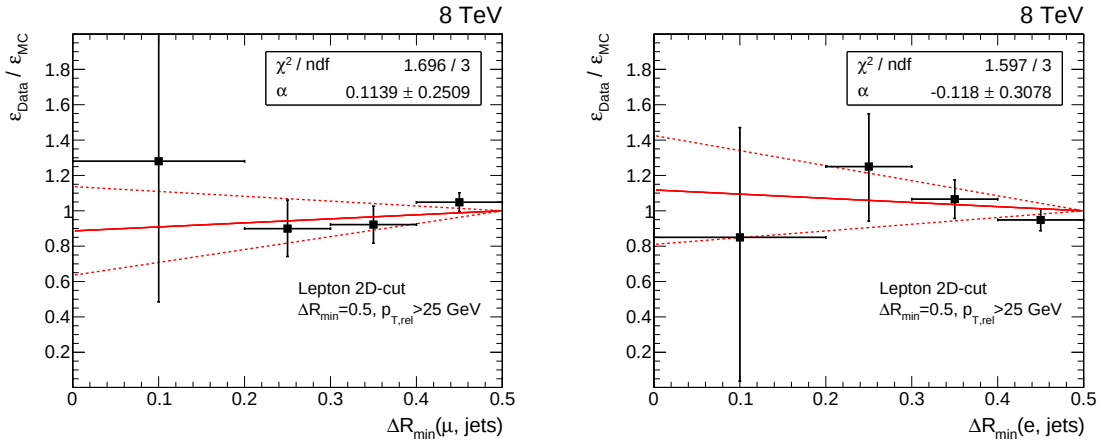


Figure 4.25: Fit of the data/MC scale factor for the lepton 2D-cut efficiency as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$ in the dimuon (left) and dielectron (right) control regions. The fit (red line) is performed with the one-parameter function given in Equation (4.7). The shaded lines correspond to the $\pm 1\sigma$ variation of the post-fit parameter (α).

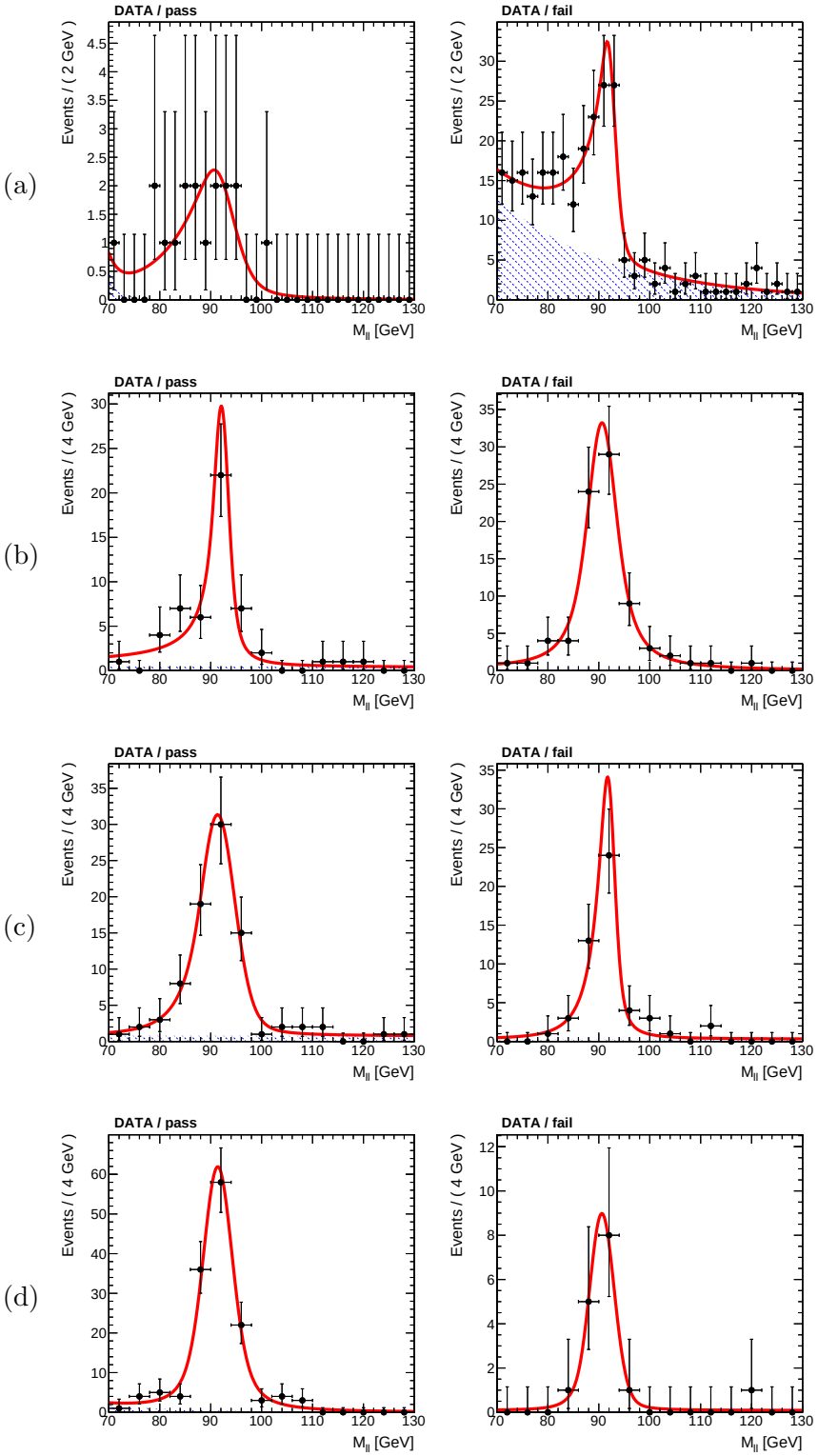


Figure 4.26: Tag-and-probe fits on $M_{\ell^+\ell^-}$ in data for the lepton 2D-cut selection in the $\mu^+\mu^-$ CR: (a) $0 < \Delta R_{\min}(\mu, j) < 0.2$, (b) $0.2 < \Delta R_{\min}(\mu, j) < 0.3$, (c) $0.3 < \Delta R_{\min}(\mu, j) < 0.4$, (d) $0.4 < \Delta R_{\min}(\mu, j) < 0.5$. The solid red line corresponds to the full fit function; the background component of the fit is given by the shaded blue area.

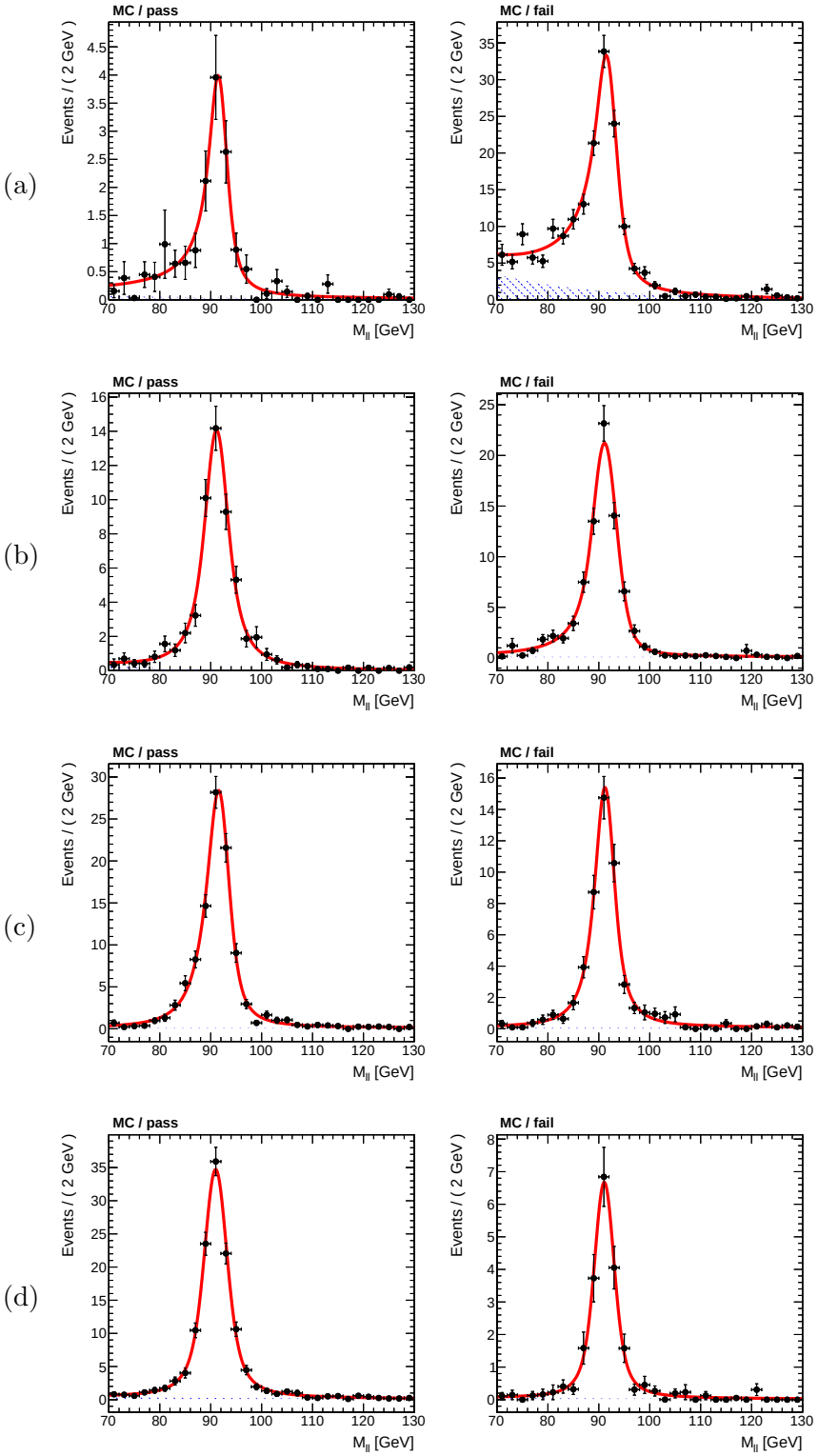


Figure 4.27: Tag-and-probe fits on $M_{\ell^+\ell^-}$ in MC for the lepton 2D-cut selection in the $\mu^+\mu^-$ CR: (a) $0 < \Delta R_{\text{min}}(\mu, j) < 0.2$, (b) $0.2 < \Delta R_{\text{min}}(\mu, j) < 0.3$, (c) $0.3 < \Delta R_{\text{min}}(\mu, j) < 0.4$, (d) $0.4 < \Delta R_{\text{min}}(\mu, j) < 0.5$. The solid red line corresponds to the full fit function; the background component of the fit is given by the shaded blue area.

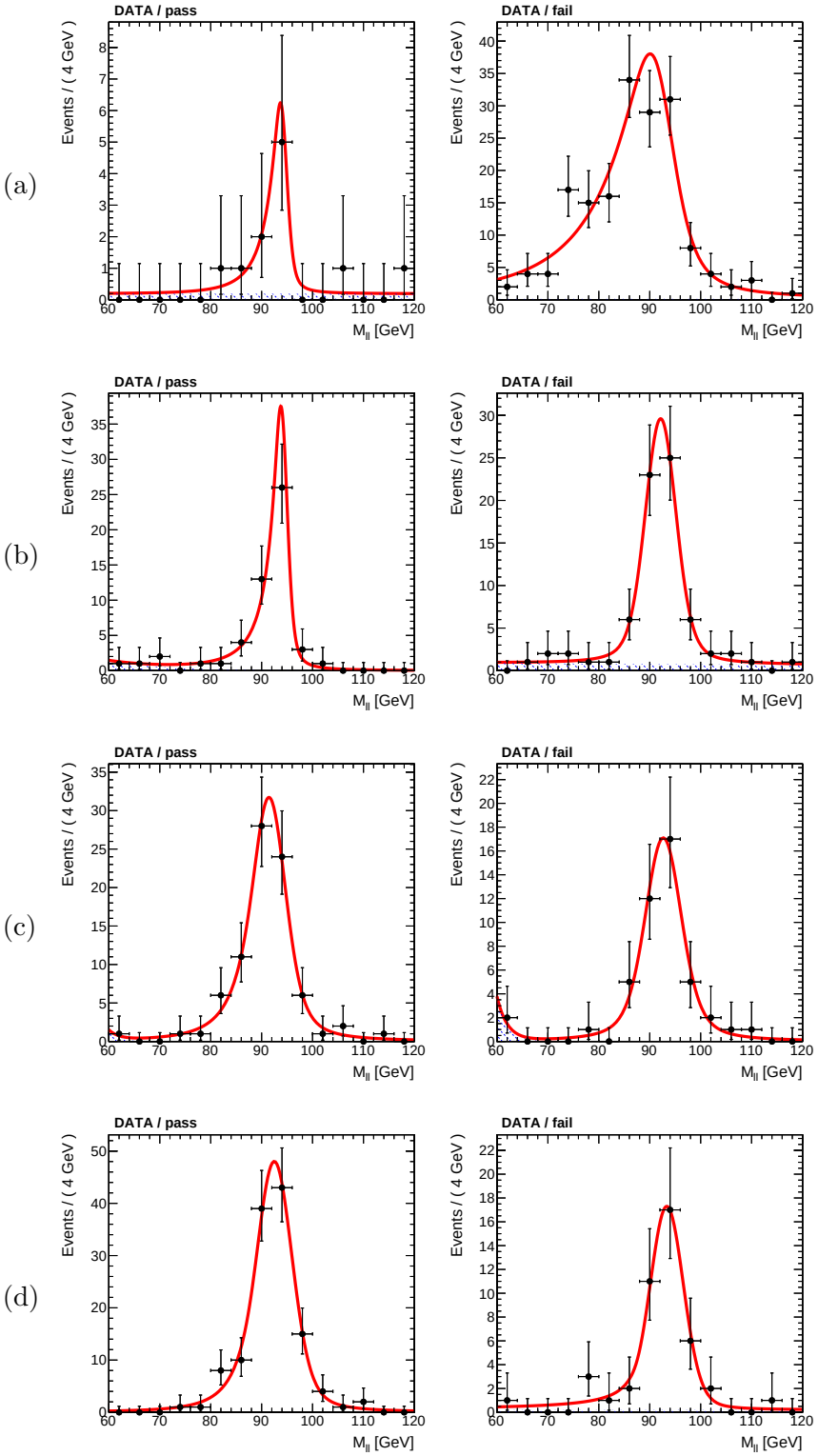


Figure 4.28: Tag-and-probe fits on $M_{\ell^+\ell^-}$ in data for the lepton 2D-cut selection in the e^+e^- CR: (a) $0 < \Delta R_{\min}(e,j) < 0.2$, (b) $0.2 < \Delta R_{\min}(e,j) < 0.3$, (c) $0.3 < \Delta R_{\min}(e,j) < 0.4$, (d) $0.4 < \Delta R_{\min}(e,j) < 0.5$. The solid red line corresponds to the full fit function; the background component of the fit is given by the shaded blue area.

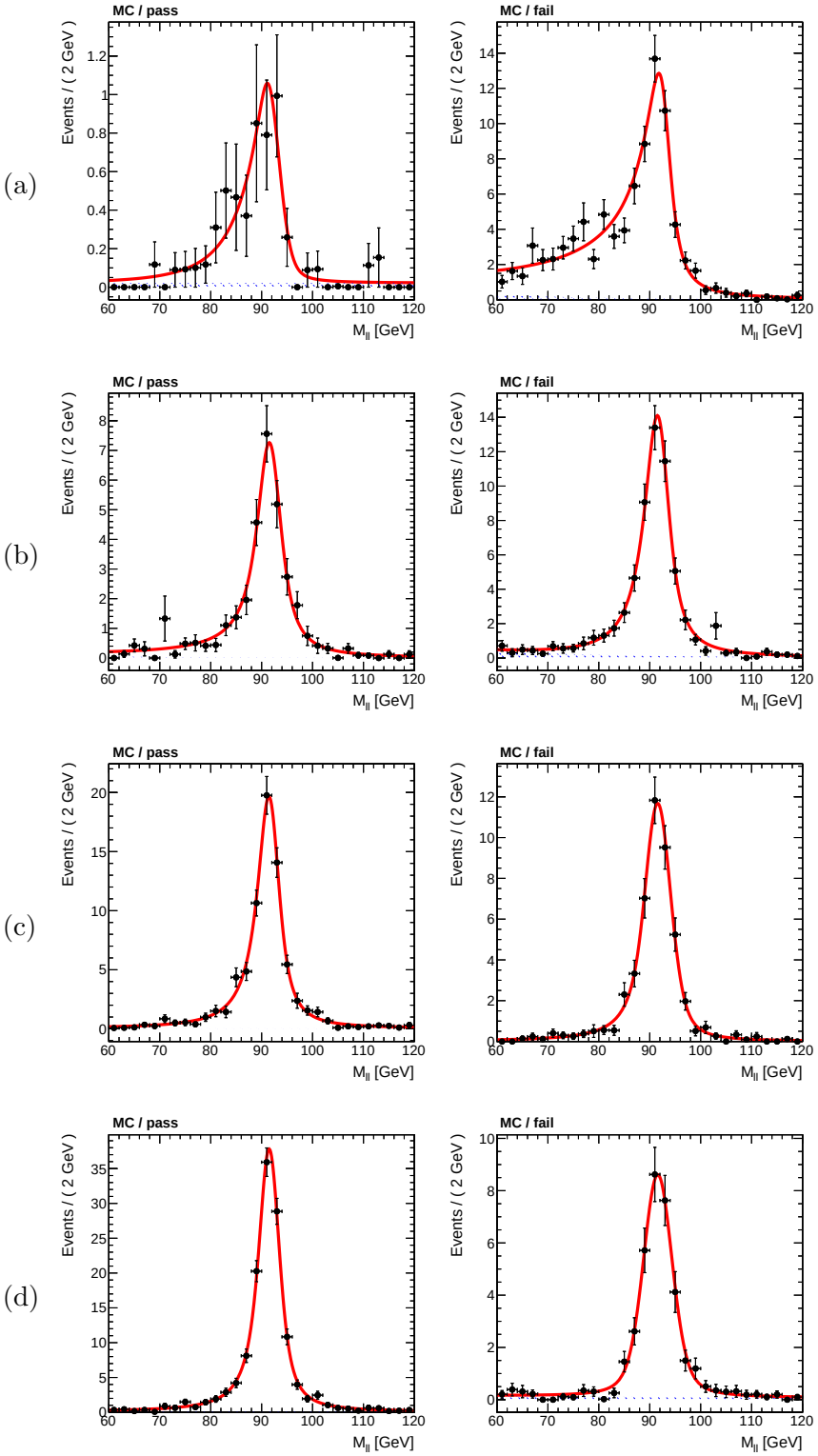


Figure 4.29: Tag-and-probe fits on $M_{\ell^+\ell^-}$ in MC for the lepton 2D-cut selection in the e^+e^- CR: (a) $0 < \Delta R_{\text{min}}(e,j) < 0.2$, (b) $0.2 < \Delta R_{\text{min}}(e,j) < 0.3$, (c) $0.3 < \Delta R_{\text{min}}(e,j) < 0.4$, (d) $0.4 < \Delta R_{\text{min}}(e,j) < 0.5$. The solid red line corresponds to the full fit function; the background component of the fit is given by the shaded blue area.

4.5.3 Jet b-tagging efficiency

As described in Section 4.4, jet b-tagging is used in the analysis to categorize the events of the final ℓ + jets SR. Indeed, the requirement of one or more b-tagged jets in the ℓ + jets final state provides a very pure sample of $t\bar{t}$ events (see Section 4.7). The specific b-tagging selection considered in the analysis for AK5 jets corresponds to the medium working point of the CSV b-tagging algorithm (CSVM), which is expected to yield a mis-identification rate of approximately 1%. Potential discrepancies between the b-tagging efficiencies in data and simulated events are accounted for according to the following procedure. Simulated events are reweighted on an event-by-event basis using the correction factor

$$w_{\text{b-tag}} = \prod_{\substack{i=1 \\ \text{(b-tagged)}}} \text{SF}_f(p_T^i\,,\,\eta^i) \prod_{\substack{j=1 \\ \text{(not b-tagged)}}} \frac{1 - \varepsilon_f^{\text{MC}}(p_T^j) \cdot \text{SF}_f(p_T^j\,,\,\eta^j)}{1 - \varepsilon_f^{\text{MC}}(p_T^j)} \tag{4.8}$$

where the first (second) product loops over the AK5 jet candidates passing (not passing) the CSVM operating point. The factor ε^{MC} corresponds to the efficiency of the CSVM algorithm in MC events, while $\text{SF} \equiv \varepsilon^{\text{data}}/\varepsilon^{\text{MC}}$ stands for the data/MC SF of a given b-tagging efficiency. Both these quantities depend on the parton-flavor f assigned to the reconstructed jet. Each jet in simulated events is associated to either a b-quark, a c-quark or, generically, a light parton (u, d, s or gluon). The jet-parton assignment is performed with a geometric criterion: the jet is matched to a generator-level parton if the latter has an angular distance $\Delta R < 0.3$ with respect to the jet axis; if more than one parton is matched to a jet, the latter is assigned the flavor of the heaviest parton (b is given priority over c, c is given priority over light and top quarks are not considered in this procedure); if the jet is not matched to a parton, it is marked as a light-flavor jet. In the following, the efficiency of the b-tagging algorithm for jets associated to a light-flavor parton is also referred to as the *b-mistag rate*. Based on the above, the numerator (denominator) in $w_{\text{b-tag}}$ corresponds to the probability of having a certain configuration of b-tagged jets in data (MC); the ratio gives the reweighting factor needed to correct the average b-tagging response in the simulation.

The b-tagging efficiencies in MC for light-, c- and b-flavor jets are determined as a function of the jet p_T for events passing the ℓ + jets SR selection. The CSVM efficiencies for the three jet flavors are shown in Figure 4.30, for the μ + jets and e + jets channels separately. For b-flavor jets, the efficiency of the CSVM selection stays between 65% and 75% for jets with moderately high p_T values (30 GeV $< p_T <$ 200 GeV) and it goes down to approximately 30% for very high-p_T jets (200 GeV $< p_T \lesssim$ 1 TeV); this reduction is due to the dependence of the heavy-hadron decay length on the jet p_T and the decreased efficiency for the reconstruction of secondary vertices inside high-p_T jets [84]. The efficiency for light-flavor jets increases as a function of the jet p_T, but it remains between 1.0% and 2.0%, as expected for the CSVM working point.

For the data/MC SFs of the b-tagging efficiencies for each of the three jet flavors, we make use of a set of independent measurements performed by the CMS Collaboration [85]; each of these data/MC corrections is measured as function of the jet p_T and η. The relative error of the efficiency SF for b-flavor jets goes from approximately 2%, for jets with 30 GeV $< p_T <$ 200 GeV, up to 5%, for higher values of the jet p_T; the uncertainty for the b-mistag rate SF ranges between 10% and 15%.

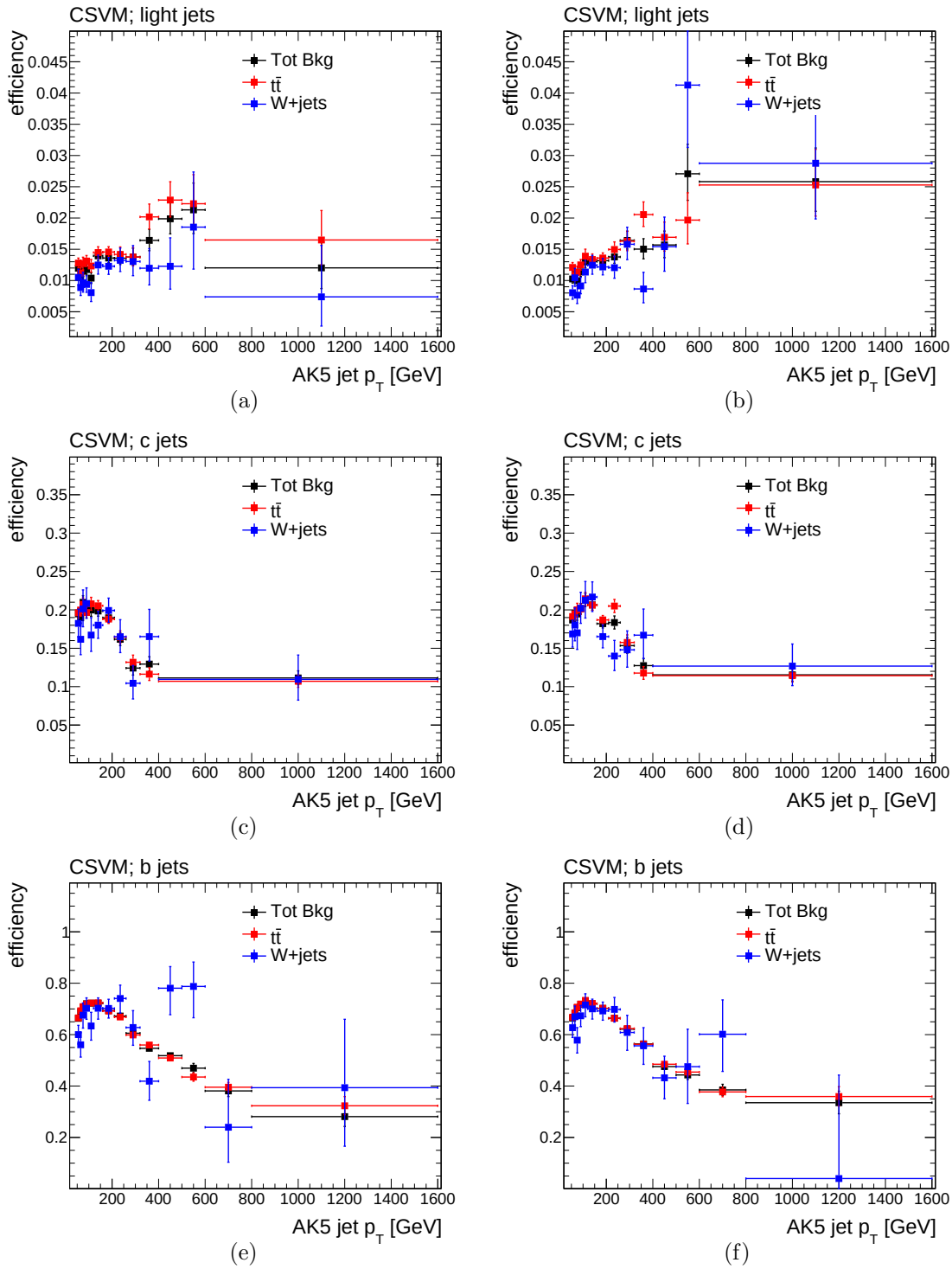


Figure 4.30: CSVM b-tagging efficiencies as a function of the jet p_T for events passing the full analysis selection described in Section 4.4. Plots are shown for the three jet flavors in the μ + jets channel (left) and in the e + jets channel (right): (a,b) light-flavor jets, (c,d) c-flavor jets, (e,f) b-flavor jets. Each plot includes the efficiency for $t\bar{t}$, W + jets and the total background.

4.5.4 Jet t-tagging efficiency

The CMS Top Tagging algorithm is used in the analysis to identify hadronically decaying top quarks reconstructed as a single large-radius jet. Like for jet b-tagging, the t-tagging method is not used to reject events, but only to define the final categories of the ℓ + jets SR.

Each CA8 jet candidate in simulated events can be assigned either top quark flavor ($f = t$) or light-parton flavor ($f = l$). A reconstructed jet is assigned top-flavor if a generated top quark is found within a cone of radius $R = 0.8$ around the jet axis; otherwise, the jet is marked as a light-flavor jet. The t-tagging efficiency for CA8 jets associated to a light parton is also referenced in the following as the *t-mistag rate*. The t-tagging efficiencies in MC are reported in Figure 4.31 for these two jet flavors and for different simulated processes; the efficiencies are measured separately in the μ + jets and e + jets channels, as a function of the CA8 jet p_T. The efficiency for true t-jets in SM $t\bar{t}$ events grows from 30% up to over 50% going from 400 GeV to 800 GeV in the jet p_T; for higher p_T values (800 GeV $< p_T \lesssim 2$ TeV), the t-tagging efficiency decreases because of the increasing difficulty to resolve the subjets of the CA8 jet candidate. For jet p_T values going from 400 GeV to 2 TeV, the t-mistag rate in $t\bar{t}$ and W + jets MC events ranges between 4% and 6%, with W + jets yielding a slightly higher t-mistag rate, compared to $t\bar{t}$ production, due to the larger fraction of quark-flavor jets (as opposed to gluon-initiated jets). A higher mis-identification rate is found for high-mass Z′ MC signals; this mainly comes from highly-boosted jets initiated by a top decay, but failing the geometric matching to the generator-level top quark. These efficiency values are consistent with those reported in a series of t-tagging studies carried out in the CMS Collaboration [87].

Simulated events are reweighted to correct the performance of the t-tagging algorithm in real data events following the same approach discussed for the b-tagging corrections in Section 4.5.3. For completeness, we report here the formula used for the t-tagging weight applied to MC events in the ℓ + jets SR:

$$w_{\text{t-tag}} = \prod_{\substack{i=1 \\ \text{(t-tagged)}}} \text{SF}_f \prod_{\substack{j=1 \\ \text{(not t-tagged)}}} \frac{1 - \varepsilon_f^{\text{MC}}(p_T^j) \cdot \text{SF}_f}{1 - \varepsilon_f^{\text{MC}}(p_T^j)} \tag{4.9}$$

where the first (second) product loops over the CA8 jet candidates passing (not passing) the t-tagging requirements; the notation in Equation (4.9) corresponds in general to the one already described for Equation (4.8). The t-tagging efficiencies in MC events are determined as a function of jet p_T, while two constant values are used for the data/MC SFs associated to the t-tagging and t-mistag rates. For the latter correction factors, the following strategy is employed in the analysis.

- The data/MC SF for the efficiency of the t-tagging selection is not measured directly, but it is introduced as a free parameter in the background model; as a consequence, its value is fitted during the limit setting procedure, as detailed in Section 4.6 and Section 4.8. This SF is taken as a global correction factor, with any assumption on its dependence on the jet kinematics.

- The t-mistag efficiency is measured in data and simulation using a μ + jets sample orthogonal to the SR selection and dominated by W+jets production, which represents the main non-top background in the ℓ + jets SR. This measurement, described in Section 4.5.4.1, is used to estimate directly the data/MC SF for the t-mistag rate.

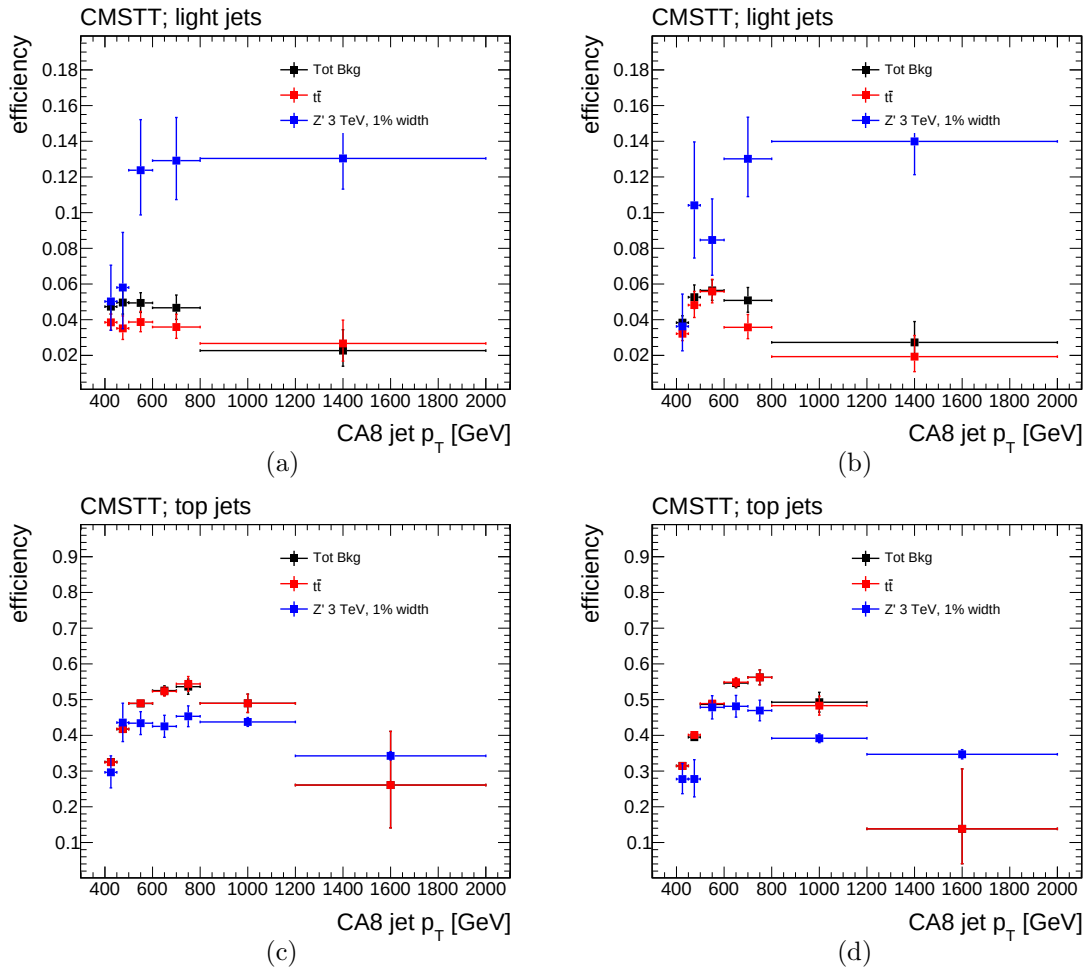


Figure 4.31: MC efficiencies for the CMS Top Tagging working point defined in Section 4.3, in events passing the full ℓ + jets selection described in Section 4.4. Efficiencies are shown as a function of the CA8 jet p_T for the μ + jets channel (left) and the e + jets channel (right): (a, b) light-flavor CA8 jets, (c, d) t-flavor CA8 jets. Each plot includes the efficiency for a 1%-width Z$'$ $\rightarrow$ signal with $M_{Z'} = 3$ TeV, SM $t\bar{t}$ production and the total SM background.

4.5.4.1 Measurement of the t-mistag efficiency in a μ + jets control sample

This section describes the measurement of the data/MC SF for the mistag efficiency of the CMS Top Tagging algorithm. The control sample used for this measurement is defined by the following selection. Each event is required to have exactly one muon candidate ($p_T > 45$ GeV and $|\eta| < 2.1$) triggering the HLT path `HLT_Mu40_eta2p1_v*` and passing the lepton 2D-cut, one CA8 jet candidate ($p_T > 400$ GeV and $|\eta| < 2.4$) and at least one AK5 jet with $p_T > 50$ GeV, $|\eta| < 2.4$ and $\Delta R > 1.3$ with respect to the CA8 jet; in addition, the event has to pass $H_{\mathrm{T}}^{\mathrm{lep}} \equiv \not\!\!E_T + p_T^{\mu} > 150$ GeV and $\not\!\!E_T > 20$ GeV. This event selection is similar to the μ + jets preselection described in Section 4.4.1 with the exception of the cut on $\not\!\!E_T$, which is relaxed from 50 GeV to 20 GeV in order to increase the statistics in this control region. The orthogonality with respect to the ℓ + jets SR is obtained by using a lower cut on the leptonic term of the χ^2 discriminator of the reconstructed $t\bar{t}$ system, given in Equation (4.4); the cut is given by $\chi^2_{\mathrm{lep}} > 50$. Cutting only on the χ^2 leptonic term avoids

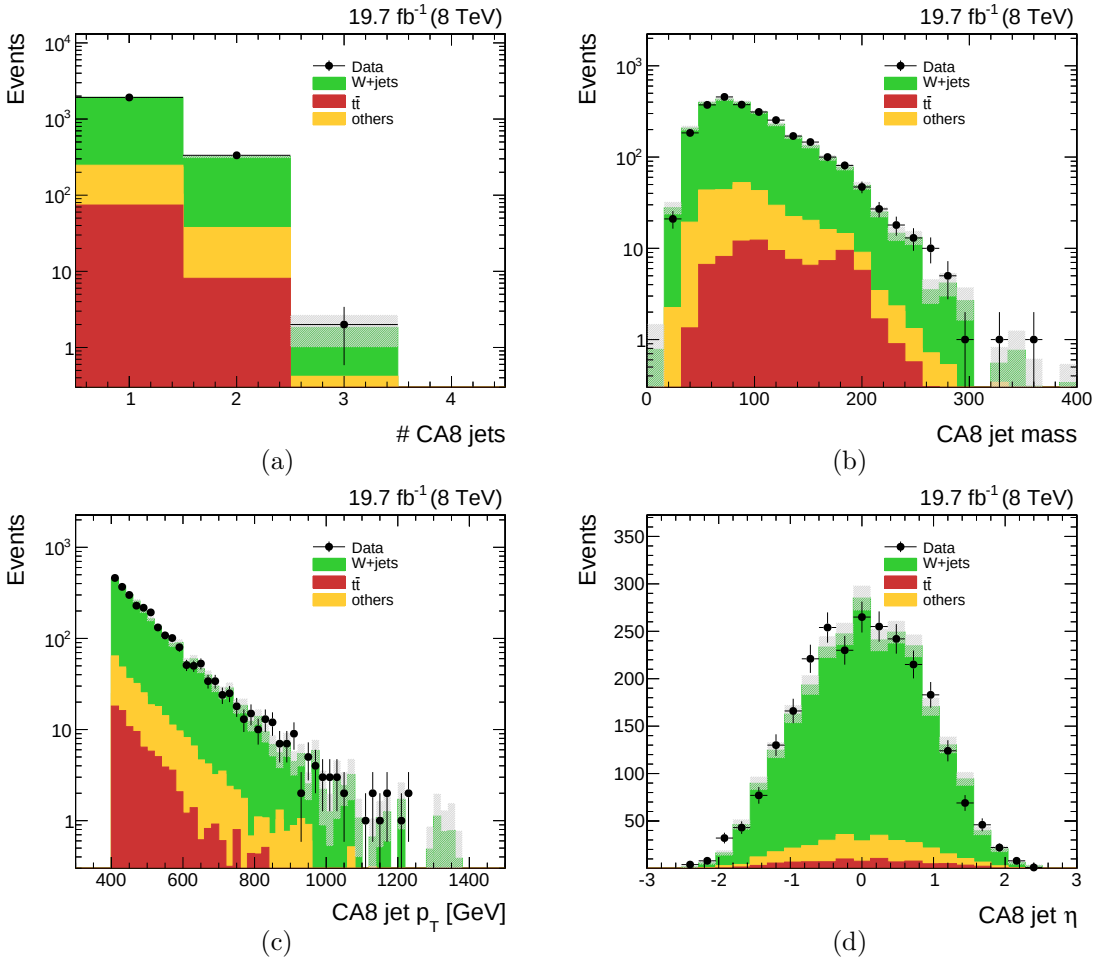


Figure 4.32: Data/MC comparisons in the control sample used for the t-mistag rate measurement (before any t-tagging requirement): (a) multiplicity of CA8 jet candidates, (b) jet mass, (c) jet p_T, (d) jet η. In each plot, the total expected background is normalized to the data. The background error band includes only the MC statistical uncertainty.

any bias on the hadronic top quark side of the event. Finally, we apply a veto on events with one or more AK5 jets passing the CSVL b-tagging selection; this is aimed at minimizing the contribution from $t\bar{t}$ events in this CR.

Figure 4.32 shows some data/MC comparisons for the kinematics of CA8 jets in the μ + jets CR defined above. The latter sample is dominated by W + jets events (89%) with a small residual contribution from $t\bar{t}$ production (less than 3%). If we consider only events in which at least one CA8 jet is t-tagged, W + jets production still accounts for the majority of the sample, but the fraction of $t\bar{t}$ events increases up to 26%. For this reason, when measuring the t-mistag rate in data, we subtract from data the contribution expected from $t\bar{t}$ production; the t-mistag efficiency in data thus corresponds to

$$\varepsilon^{\text{mistag}}_{\text{Data}} = \frac{N^{\text{tagged}}_{\text{Data}} - N^{\text{tagged}}_{t\bar{t}}}{N_{\text{Data}} - N_{t\bar{t}}} \tag{4.10}$$

where N_{Data} ($N_{t\bar{t}}$) denotes the number of CA8 jet candidates in the control sample for data ($t\bar{t}$ MC) and $N^{\text{tagged}}_{\text{Data}}$ ($N^{\text{tagged}}_{t\bar{t}}$) corresponds to the number of those CA8 jets that are t-tagged

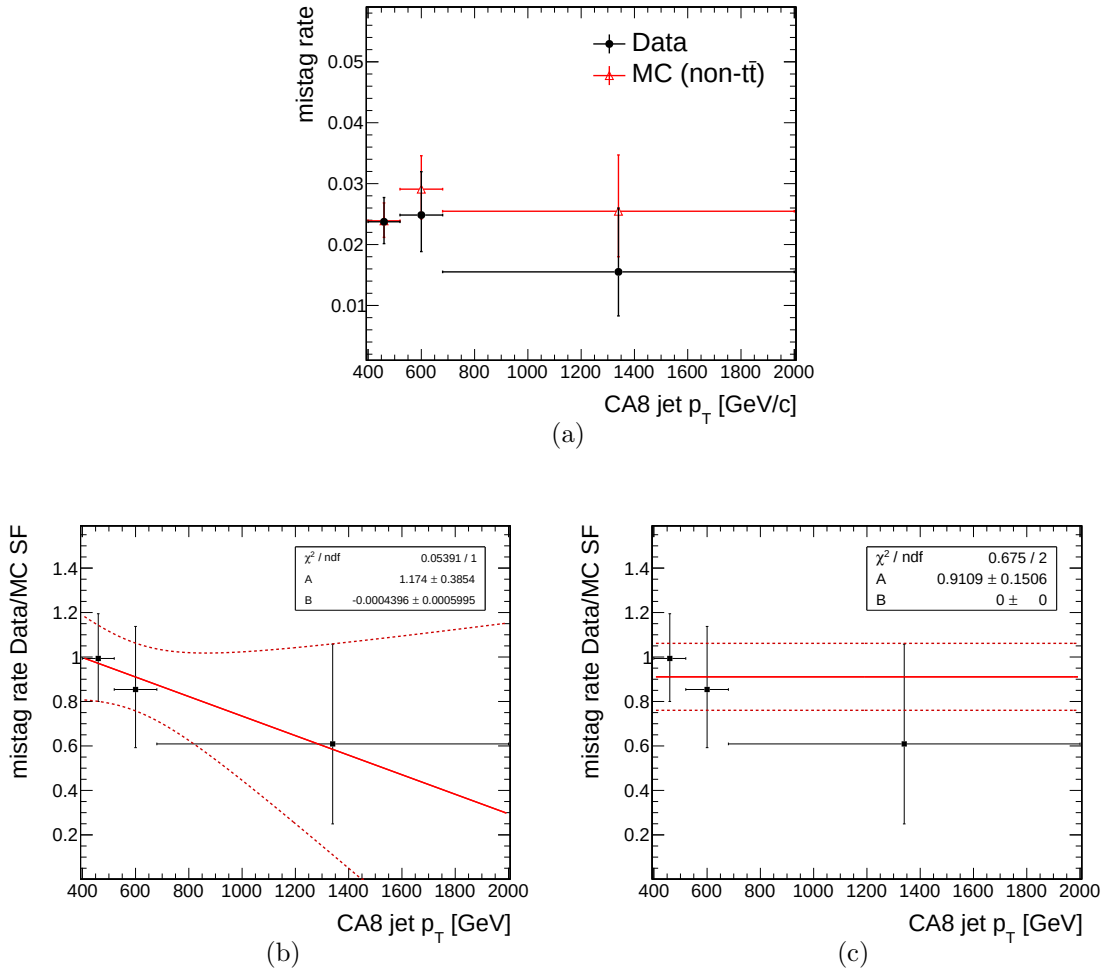


Figure 4.33: Efficiency and data/MC SF for jet t-tagging in the W($\to \mu\nu$) + jets CR: (a) t-mistag rate in data and MC as a function of the jet p_T, (b,c) data/MC t-mistag rate SF as a function of the jet p_T. The SF fit function in (b) is performed with a linear function ($f = A + B \cdot p_T$), whereas a constant function is used in (c).

in data ($t\bar{t}$ MC). The t-mistag rate in simulated events is measured directly from the W+jets MC, so we have $\varepsilon^{\text{mistag}}_{\text{MC}} = N^{\text{tagged}}_{\text{W+jets}}/N_{\text{W+jets}}$, where $N_{\text{W+jets}}$ is the number of CA8 jets in the μ + jets CR and $N^{\text{tagged}}_{\text{W+jets}}$ is the number of jets passing the t-tagging requirements. As a result, the data/MC SF for the t-mistag efficiency is given by $\varepsilon^{\text{mistag}}_{\text{Data}}/\varepsilon^{\text{mistag}}_{\text{MC}}$.

Figure 4.33 shows the t-mistag efficiencies in data and simulation, plus the corresponding data/MC scale factor, as a function of the jet p_T. After the t-tagging requirement, the statistics in the control sample is limited for both data and MC. Even using only three bins in the jet p_T spectrum, the fit of the t-mistag rate SF with a linear function, shown in Figure 4.33(b), gives very large uncertainties and the -1σ band goes down to zero in the high-p_T region. Since there is not enough statistics for a reasonable estimate of the t-mistag rate SF as a function of the jet kinematics, we restrict ourselves to the measurement of the overall t-mistag rate in data and MC and we derive one global correction factor; this is shown in Figure 4.33(c). As no prior knowledge on the SF shape is assumed, we multiply the uncertainty of the this flat correction by a factor 1.5, in order to have a conservative estimate. The final value of the t-mistag rate SF thus corresponds to $\text{SF}_l^{\text{t-tag}} = 0.91 \pm 0.23$.

4.6 Systematic uncertainties

Several sources of systematic uncertainties affect the MC predictions for background and signal processes in the ℓ + jets SR. In the context of this search, it is particularly important to quantify the impact of various systematic uncertainties on the invariant mass spectrum of the $t\bar{t}$ system, which is the most important kinematic distribution in the analysis.

A summary of the effect of each systematic uncertainty on the normalization of the $M_{t\bar{t}}$ distributions in the ℓ + jets SR is reported in Table 4.3. The uncertainties on the cross sections of SM backgrounds, the total integrated luminosity and the electron trigger efficiency act only as normalization errors; in general, though, all the other systematic uncertainties can affect both the normalization and the shape of the $M_{t\bar{t}}$ distributions; the size of a given uncertainty can vary depending on the process or event category under consideration. The main systematic errors in this analysis are due to the data/MC corrections for the jet b-tagging and t-tagging efficiencies, the uncertainties on the cross sections of the SM backgrounds and the theoretical uncertainties related to the choice of the renormalization and factorization scales and the PDF set utilized in the simulation of background and signal processes.

The $\pm 1\sigma$ $M_{t\bar{t}}$ templates for some of the most relevant systematic uncertainties can be found in Figures 4.34–4.36. In the rest of this section, the systematic uncertainties included in the analysis are described in full detail.

Cross sections of SM processes

We assign uncertainties on the inclusive cross sections of each SM process included in the background model; these uncertainties are estimated based on independent CMS measurements available in the literature at the time of this analysis. The cross section for $t\bar{t}$ production in the boosted regime is assigned an overall uncertainty of 15% [121, 122]. For W + jets production, separate uncertainties are assigned to the cross sections for W+light, W+charm and W + bottom processes: we take an uncertainty of 9% for the production of a W boson in association with light-flavor partons [123], and a 23% error for both the W + charm and W + bottom cross sections [124]. The MC prediction for the three process is determined by categorizing the events in the W + jets simulation based on the flavor content of the partons produced in association with the W boson. The single-top background in the analysis mostly originates from tW production, whose cross section has been measured to an accuracy of 23% [125]; this value is used as the uncertainty for the total single-top cross section. The cross section of Z + jets production has been measured to an accuracy of 7% and 9% for events with two and three additional jets [126]; on the other hand, since this analysis explores a rather different phase space compared to these reference measurements, we assign a 50% uncertainty to the normalization of this very small background. This conservative estimate is also intended to cover the uncertainties on the choice of the renormalization and PS-matching scales in the Z + jets simulation. The uncertainty on diboson production is taken to be 20% [127, 128].

Pileup

MC samples are reweighted so that the number of true pileup interactions in MC matches the instantaneous luminosity profile in data with a minimum-bias cross section of 69.4 mb. The systematic associated to this correction is quantified by varying the minimum-bias cross section by ±5%.

Integrated luminosity

The uncertainty on the total integrated luminosity recorded by CMS in 2012 at $\sqrt{s} = 8$ TeV is 2.6% [129].

Lepton ID and trigger efficiency

Scale factors for the efficiency of the muon ID and trigger selections are used to reweight simulated events in the ℓ + jets analysis, as a function of the muon p_T and η; similarly, data/MC SF for the electron ID efficiency are applied as a function of the electron p_T and η. These corrections have been measured with a tag-and-probe method in control samples dominated by $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events [75, 77]. The corresponding systematic uncertainty is obtained by varying each scale factor by its $\pm 1\sigma$ error. For the muon ID (trigger) SF, an additional uncertainty of 0.5% (0.2%) is added in quadrature to the statistical error of the SF in order to account for systematic uncertainties on the tag-and-probe method used to derive these corrections.

For the efficiency of the logical OR of the "electron+2-jets" and single-jet triggers, an uncertainty of 1% on the scale factor is used, based on the measurement in the $e\mu$ control sample discussed in Section 4.5.1.

Lepton 2D-cut efficiency

The efficiency of the lepton 2D-cut in simulated events is corrected using the data/MC SFs measured in the $Z \to \ell^+\ell^-$ CRs, as described in Section 4.5.2. The corresponding systematic uncertainties are determined by varying the results of the SF fit within its $\pm 1\sigma$ band. The upward (downward) variation $\alpha_{\text{fit}} + 1\sigma$ ($\alpha_{\text{fit}} - 1\sigma$) corresponds to a lower (higher) lepton 2D-cut efficiency; as this systematics acts only on events in the region $\Delta R_{\text{min}}(\ell, \text{jets}) < 0.5$, it affects the normalization of the highest-mass signal hypotheses and it has a limited impact on the SM backgrounds (below 2% overall).

Jet Energy Scale

The systematic uncertainty due to jet energy scale corrections applied on AK5 and CA8 jets is evaluated by varying the JEC applied on each jet by its $\pm 1\sigma$ error [114, 115]. The variation in the energy scales of AK5 jets is also propagated to the $\not\!\!E_T$ measurement. For the jet candidates considered in the analysis ($p_T > 30$ GeV, $|\eta| < 2.4$), the total uncertainties on the JECs are typically smaller than 3%; in particular, the uncertainties for jets in the barrel region ($|\eta| < 1.3$) are below 1%. The JEC uncertainties for AK5 and CA8 jets are treated as fully correlated; this is motivated by the fact that the two sets of JECs are determined using the same methodology.

Jet Energy Resolution

The jet energy resolution of AK5 and CA8 jets in simulated events is corrected to match the energy resolution measured in data [114, 116]; the JER scale factors depend on the jet η and their uncertainty goes from 2% for $|\eta^j| < 0.5$ to 4% for $|\eta^j| < 2.3$. The corresponding systematic uncertainty is measured by varying these scale factors by their $\pm 1\sigma$ error. Similarly to the JEC uncertainty, the effect of the JER variations is propagated to the $\not\!\!E_T$ measurement. The same scale factors are used for AK5 and CA8 jets and this systematic is treated as fully correlated among the two jet collections.

b-tagging and b-mistag efficiencies

The b-tagging efficiencies in MC events are corrected according to the procedure discussed in Section 4.5.3. The systematic uncertainty related to the b-tagging correction for each jet flavor ($f = l, c, b$) is derived by propagating the $\pm 1\sigma$ error of the corresponding data/MC SF in the reweighting factor of Equation (4.8). The uncertainties on SF_b and SF_c are treated as fully correlated, while the systematic error due to the b-mistag rate correction SF_l is taken as uncorrelated from the first two.

t-tagging and t-mistag efficiencies

The systematic uncertainties of the MC prediction related to the efficiency and mistag rate of the t-tagging selection are determined by varying the corresponding data/MC SFs in the reweighting factor of Equation (4.9). The data/MC SF for the tagging efficiency on true t-jets is not measured directly, but it is introduced in the analysis as a free parameter and constrained *in situ* during the limit setting procedure. For the t-mistag data/MC, we use the value measured in a sideband region dominated by W + jets production, as detailed in Section 4.5.4.1; the uncertainty of this SF corresponds to 25%. The uncertainties on the t-tagging and t-mistag efficiency corrections are treated as uncorrelated.

Parton Distribution Functions (PDFs)

The systematic error associated to the PDF choice in each signal and background process is determined by reweighting the original MC sample independently for each eigenvector of its PDF set [130]. The variations determined for each individual eigenvector are added in quadrature in each bin of a given distribution in order to derive the $\pm 1\sigma$ systematic templates.

Renormalization and factorization scale ($t\bar{t}$ and W + jets)

We study the effect of the choice of the renormalization and factorization Q^2-scales in the MC simulation for the $t\bar{t}$ and W + jets backgrounds. This systematic error is estimated using independent MC samples generated with scales equal to $\mu_{\mathrm{down}} = Q^2/4$ and $\mu_{\mathrm{up}} = 4Q^2$.

Parton shower matching scale (W + jets)

Similarly to what is done for the Q^2-scale systematic, an additional uncertainty is assigned to the W + jets sample for the choice of the matching scale, $Q_{\text{ME-PS}}$, between the matrix element (ME) and parton-shower (PS) calculations in the MC simulation. Dedicated MC samples, with scales equal to $Q^{\mathrm{up}}_{\text{ME-PS}} = 2Q_{\text{ME-PS}}$ and $Q^{\mathrm{down}}_{\text{ME-PS}} = Q_{\text{ME-PS}}/2$ are used to quantify the size of this systematic. The typical uncertainty for the ME-PS matching scale is typically around 3% and it is smaller that the systematic error due to the choice of the Q^2-scale.

Table 4.3: Systematic uncertainties on the normalization of the $M_{t\bar{t}}$ distributions for different background and signal processes. The values in the table correspond to the percentage difference with respect to the nominal MC yield. When necessary, the effect of a given uncertainty is reported separately for different event categories. Squared brackets, when present, denote the range of a given uncertainty across different categories or processes. The +/− symbols, when present, indicate the correlation between the systematic uncertainty and the variation of the MC yield. For the t-tagging efficiency, uncertainties are quoted for an arbitrary variation of 20% of the data/MC SF, only for illustration purposes.

Systematic uncertainty	Background	Z′ signal (1% width)		
		$M = 1$ TeV	$M = 2$ TeV	$M = 3$ TeV
SM cross sections	15% ($t\bar{t}$) 9% (W + light) 23% (W + c) 23% (W + b) 50% (Z + jets) 23% (single-top) 20% (diboson)			
luminosity	+2.6%			
pileup	0.7%	1.5%		
muon ID and trigger efficiencies	[+1.3%, +2.0%]	+1.7%	+2.3%	+2.7%
electron ID efficiency	+0.5%	0.7%		
electron trigger efficiency	+1%			
lepton 2D-cut efficiency	[−0.6%, −1.5%]	−0.8%	−3.9%	−6.8%
JEC	[+2.5%, +6.8%]	+2.7%	1.4%	1.1%
JER	[+1.3%, +2.1%]	[+0.4%, +1.8%]		
b-mistag rate (`T0B0`)	−0.6%	−0.3%		
b-mistag rate (`T0B1`)	+0.2% ($t\bar{t}$) , +4.0% (W + jets)	[+0.2%, +0.4%]		
b-mistag rate (`T1`)	+0.2%	+0.1%		
b-tagging efficiency (`T0B0`)	−4.6% ($t\bar{t}$) , −0.3% (W + jets)	−6.3%	−7.6%	−8.6%
b-tagging efficiency (`T0B1`)	+1.3% ($t\bar{t}$) , +2.0% (W + jets)	+1.3%	+2.4%	+3.0%
b-tagging efficiency (`T1`)	+0.2%	−0.2%	−1.0%	−2.1%
t-mistag rate (`T0B0`, `T0B1`)	−0.1%	−0.1%		
t-mistag rate (`T1`)	+1.7% ($t\bar{t}$) , +25.0% (W + jets)	+0.2%	+0.7%	+1.2%
t-tagging efficiency (`T0B0`, `T0B1`)	−1.6% ($t\bar{t}$)	−17.0%	−15.0%	−14.4%
t-tagging efficiency (`T1`)	+18.8% ($t\bar{t}$)	+19.8%	+19.4%	+19.0%
ME Q^2-scales, $t\bar{t}$ (`T0B0`, `T0B1`)	[6.9%, 11.4%]			
ME Q^2-scales, $t\bar{t}$ (`T1`)	14.5%			
ME Q^2-scales, W + jets	[13.1%, 15.2%]			
PS matching scale (W + jets)	[1.5%, 3.2%]			
PDFs	[11.0%, 16.1%] ($t\bar{t}$) [4.0%, 9.8%] (W + jets)	+2.0%	+3.0%	+7.3%

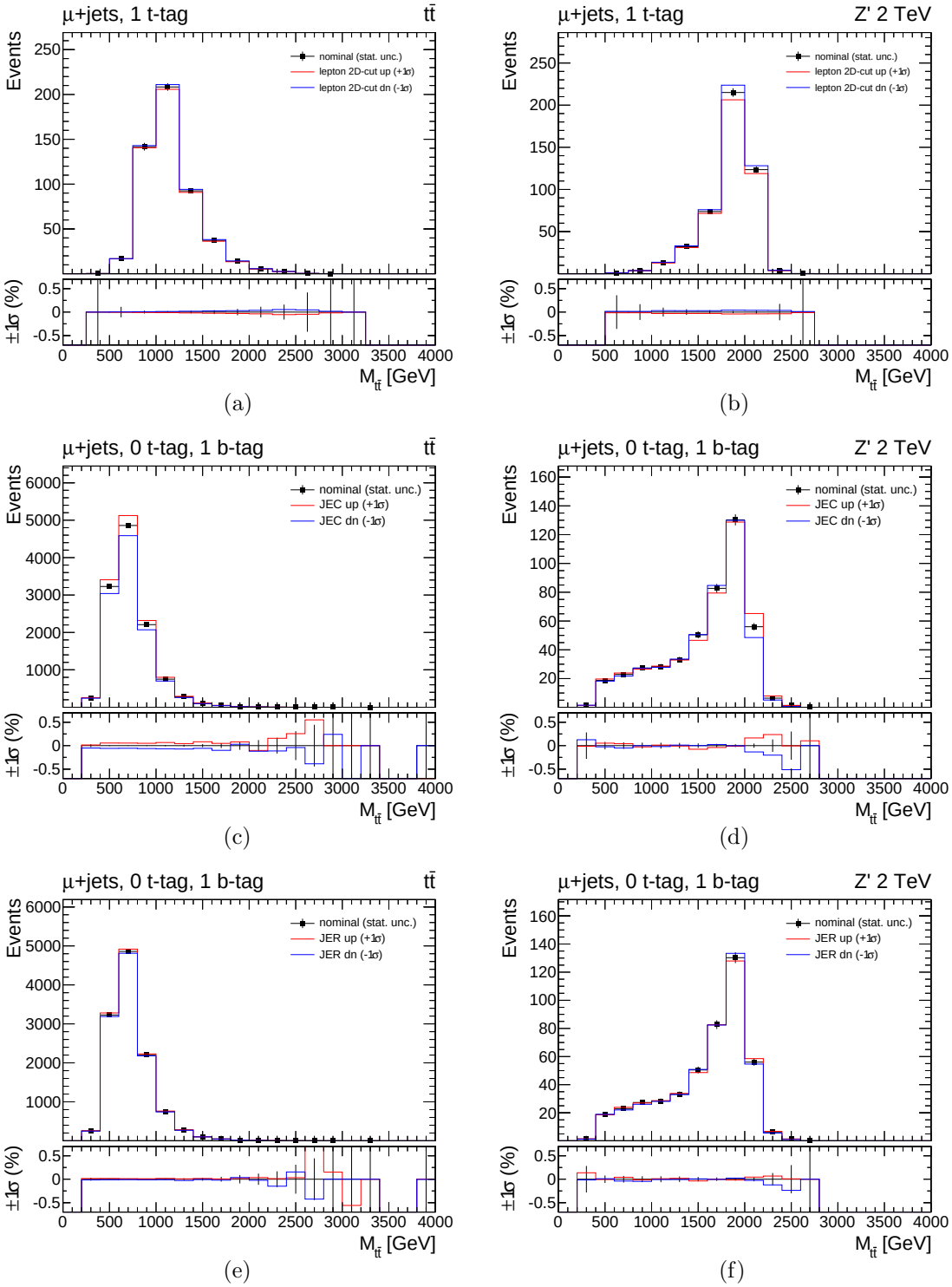


Figure 4.34: Systematic uncertainties on the $M_{t\bar{t}}$ distributions in the μ+jets channel for SM $t\bar{t}$ production (left) and a Z′ signal with $M_{Z'} = 2$ TeV and 1% width (right): (a, b) lepton 2D-cut efficiency SF in the `T1` category, (c, d) JEC in in the `T0B1` category, (e, f) JER in in the `T0B1` category.

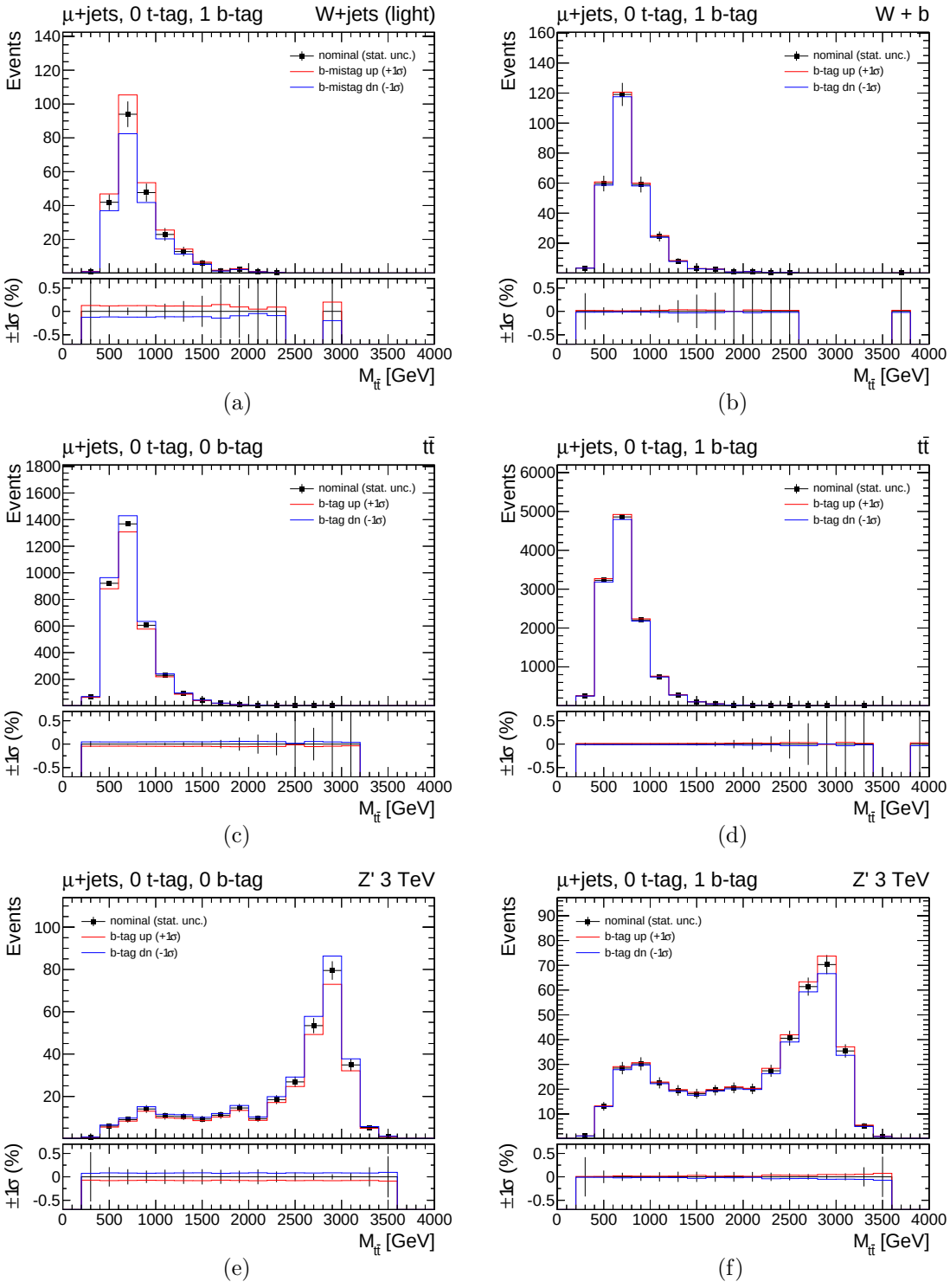


Figure 4.35: Systematic uncertainties on the $M_{t\bar{t}}$ distributions in the μ + jets channel: (a) b-mistag rate correction for W + light production in the `T0B1` category, (b) b-tagging efficiency correction for W + b production in the `T0B1` category, (c, d) b-tagging efficiency correction for $t\bar{t}$ production in the `T0B0` and `T0B1` categories, (e, f) b-tagging efficiency correction in the `T0B0` and `T0B1` categories for a 1%-width Z′ with $M_{Z'} = 3$ TeV.

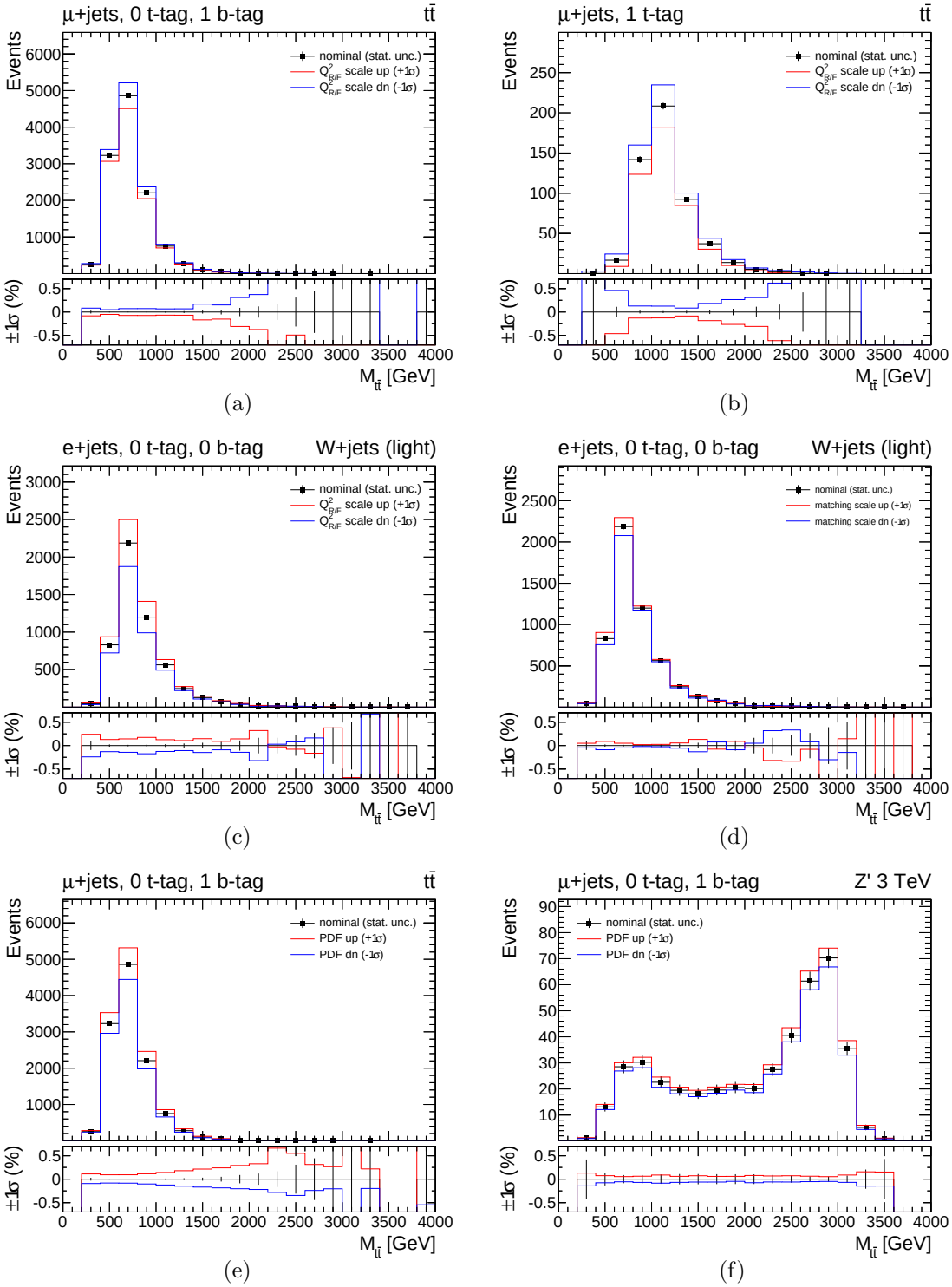


Figure 4.36: Systematic uncertainties on the $M_{t\bar{t}}$ distributions related to theoretical uncertainties in the MC simulation: (a, b) Q^2-scale uncertainty for $t\bar{t}$ production in the `T0B1` and `T1` categories, (c, d) Q^2-scale and PS-matching scale uncertainties for W + light production in the `T0B0` category, (e, f) PDF uncertainty in the `T0B1` category for $t\bar{t}$ production and a 1%-width Z′ with $M_{Z'} = 3$ TeV.

4.7 Background model

All SM backgrounds in the analysis are estimated from MC simulations. As described in previous sections, every MC sample is reweighted in order to account for known discrepancies between data and simulated events, for example regarding the pileup multiplicity, trigger efficiency, lepton ID and jet tagging efficiencies. For a detailed discussion of all these corrections and their uncertainties we refer to Section 4.6.

The expected yield of the simulated SM processes in the different ℓ+jets SR categories is obtained by means of a binned maximum-likelihood fit of the background-only model to the data. This procedure is introduced to correct the central value of the MC background prediction (within its uncertainties) in the SR categories, but it is not used to modify the size of the systematic uncertainties associated to the background and signal MC distributions. The likelihood minimization is performed using the THETA software [131]. The observables used in the fit are the reconstructed $M_{t\bar{t}}$ distributions in the six categories of the ℓ+jets SR, which are fitted simultaneously. The systematic uncertainties listed in Section 4.6 are introduced as nuisance parameters of the background-only fit. As already mentioned in Section 4.5.4, the t-tagging SF is included in the likelihood fit as a free parameter. For the other nuisance parameters in the likelihood, log-normal prior distributions are assigned to systematic uncertainties affecting only the normalization of a given process, while a gaussian prior is used for each of the systematic uncertainties that affect the shape of the $M_{t\bar{t}}$ distributions.

The post-fit values for the background-only model are shown in Figure 4.37. The post-fit nuisance parameters are expressed in units of their prior uncertainty. This means that, for a given nuisance parameter, a post-fit value of -0.2 ± 0.4 indicates that the fit result lies at $-0.2 \cdot \sigma$ with respect to the pre-fit mean value and the post-fit uncertainty equals $0.4 \cdot \sigma$, where σ stands for the pre-fit uncertainty assigned to the nuisance parameter. The fit has also been repeated with the injection of different signal hypotheses, in order to verify the stability of the post-fit values obtained using the background-only model; the corresponding results are compatible with the ones shown in Figure 4.37. None of the post-fit values lies outside of the 2σ band and none of the post-fit uncertainties departs considerably from its original value: this suggests that the initial estimate of each nuisance parameter is in reasonable agreement with the fit results. The fit results also show which are the parameters that can be best constrained by the present analysis. Examples of this are the $t\bar{t}$ cross section, the t-tagging SF and the theoretical uncertainties on the $t\bar{t}$ and W+jets backgrounds: the post-fit value of the $t\bar{t}$ cross section and t-tagging SF are driven by the background normalization in the `T0B1` and `T1` categories, while the systematics due to the theoretical uncertainties on the $t\bar{t}$ and W+jets backgrounds are constrained by the shape of the $M_{t\bar{t}}$ distributions in the different SR categories. On the other hand, as mentioned above, the post-fit uncertainties derived from this likelihood fit are not used further in the analysis. For the limit-setting procedure, we use the prior uncertainties described in Section 4.6 and leave the t-tagging SF unconstrained; we only apply the post-fit mean values shown in Figure 4.37 to correct the central values of the nuisance parameters of the analysis, in order to improve the overall agreement between the data and the background model. Figures 4.38–4.39 show a comparison between the $M_{t\bar{t}}$ distributions in data and the ones predicted for the total background, before and after the likelihood fit described above.

The overall event yields expected from SM processes (after the background-only fit) and observed in data are reported in Table 4.4 for each of the SR categories; the observed and expected number of events are in agreement with each other, within their uncertainties, in each of the six event categories. As shown in these tables, the two main SM backgrounds

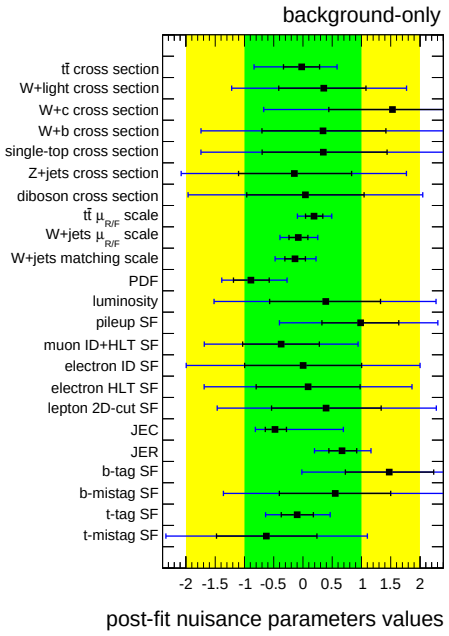


Figure 4.37: Post-fit values of the nuisance parameters of the background model, obtained from a binned maximum-likelihood fit to the data using the $M_{t\bar{t}}$ distributions. Values are shown for the background-only hypothesis. The deviation of each nuisance parameter from its pre-fit value is expressed in units of its prior uncertainty; the $\pm 1\sigma$ ($\pm 2\sigma$) post-fit error band for each parameter is shown in black (blue). The t-tagging SF parameter is left unconstrained in the fit and its post-fit value is shown in units of a prior uncertainty of 20%.

in ℓ + jets SR are given by $t\bar{t}$ and W + jets production: the latter accounts for a sizeable portion of the total background in the `T0B0` category, while the former fully dominates the `T0B1` and `T1` categories. The overall contribution of other SM backgrounds, namely single-top, Z + jets and diboson production, is expected to be rather small in all categories. Figures 4.40–4.41 show the corresponding data/MC comparison for the mass distributions of the reconstructed $t\bar{t}$ pair; in these plots, the error band assigned to the background prediction includes the MC statistical uncertainty and the post-fit systematic errors as determined from the background-only fit described above. We do not observe any significant excess in the $M_{t\bar{t}}$ distributions which would indicate the presence of resonant $t\bar{t}$ production in the data. A small overflow of events in the data is present in the mass region between 1.2 TeV and 1.4 TeV for the `T1` category of the μ+jets analysis and the `T0B1` category of the e+jets analysis. On the other hand, no sign of a similar excess is observed in any of the other four SR categories; in particular, neither excess shows the expected consistency across the μ+jets and e+jets. This suggests that these discrepancies are only the effect of statistical fluctuations. In general, the observed and expected $M_{t\bar{t}}$ distributions are in agreement within 1σ (2σ), when considering the total pre-fit (post-fit) uncertainty associated to the background prediction. Aside from the excesses discussed above, good agreement is observed in the $M_{t\bar{t}}$ spectra between the data and the expected SM background.

In order to test the reliability of the background model, we also verify the agreement between data and MC for a number of additional kinematic quantities relevant to the analysis. Several of these distributions are shown in Figures 4.42–4.45 for events in the ℓ + jets SR. In general, good agreement is found between the data and the background-only model for all the distributions considered.

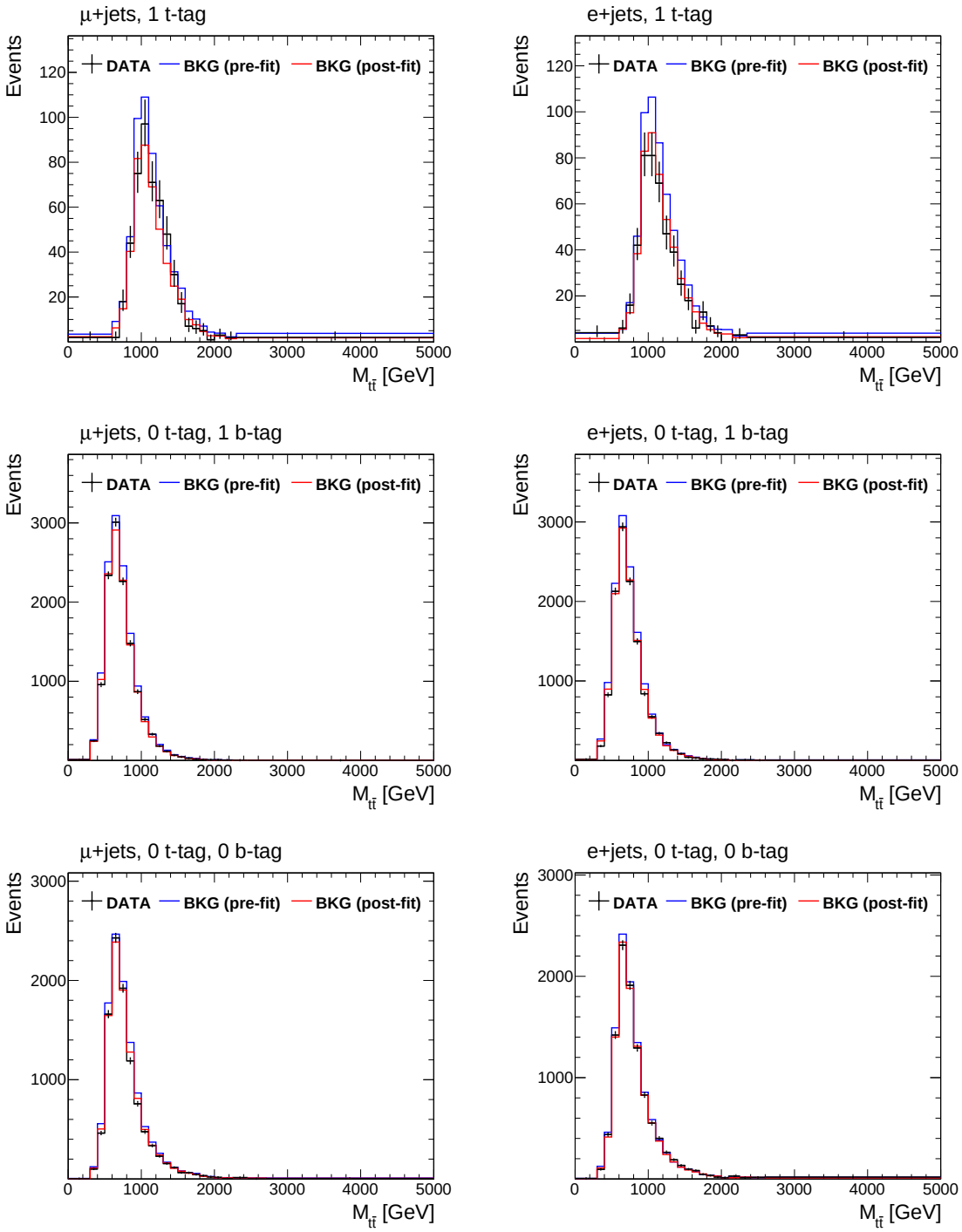


Figure 4.38: Comparison of the $M_{t\bar{t}}$ spectra in data (black) with the expected distributions of the background model, before (blue) and after (red) the maximum-likelihood fit described in Section 4.7. Each plot corresponds to one of the six SR categories in the ℓ + jets analysis.

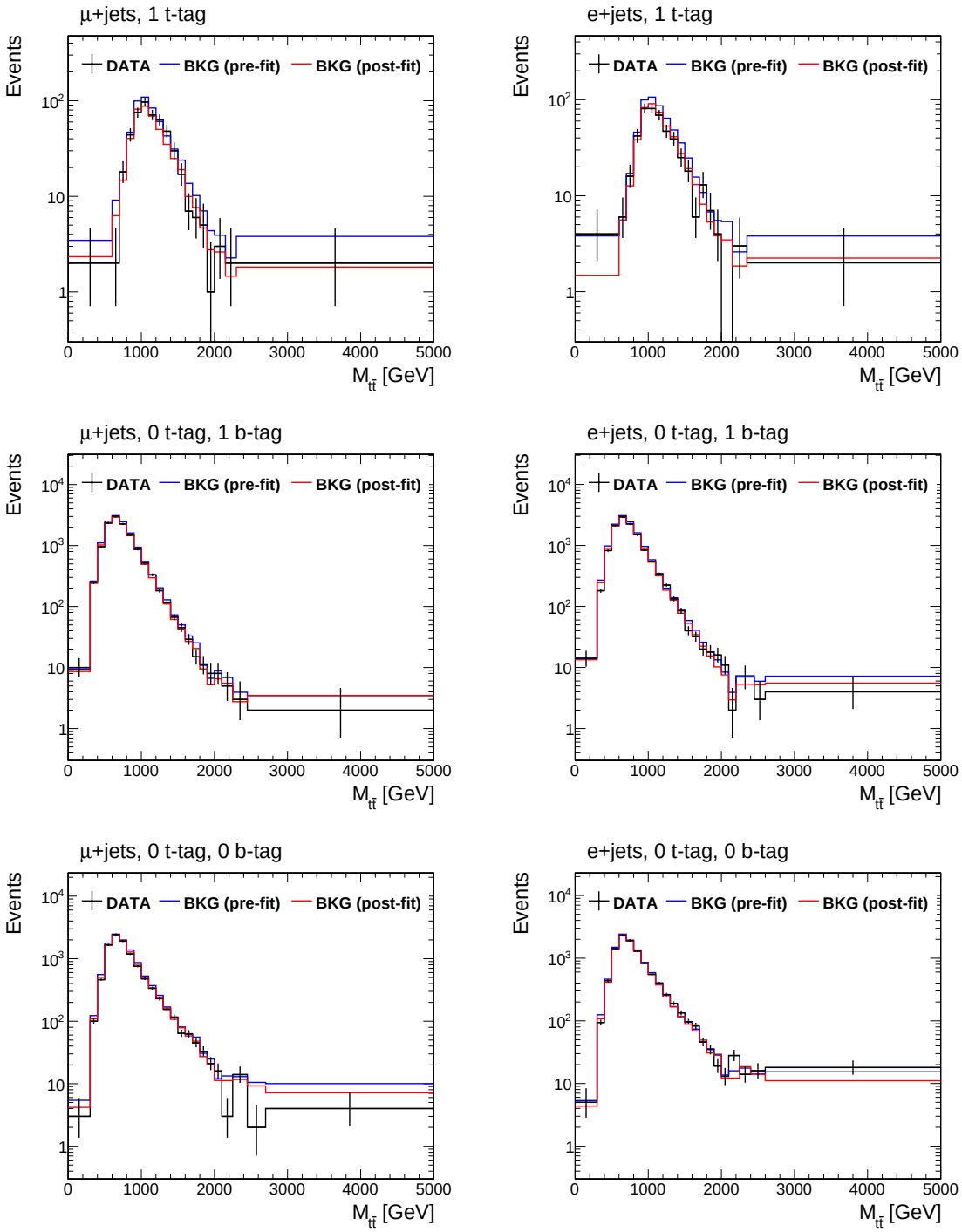


Figure 4.39: Same distributions of Figure 4.38, shown in logarithmic scale.

Table 4.4: Number of observed and expected events in the six SR categories of the ℓ + jets analysis. The expected yields for SM processes are obtained from the background-only fit to the data described in the text. The error associated to the SM backgrounds includes the MC statistical uncertainty and all the post-fit systematic uncertainties.

μ + jets SR

Process	T1	T0B1	T0B0
$t\bar{t}$	418 ± 39	10640 ± 694	2717 ± 204
W + jets (light)	23 ± 6	235 ± 35	5255 ± 422
W + jets (c, b)	13 ± 5	752 ± 153	1647 ± 432
single-top + DY + VV	9 ± 3	777 ± 185	655 ± 181
Total Background	462 ± 42	12404 ± 779	10275 ± 732
Data	493	12510	10099

e + jets SR

Process	T1	T0B1	T0B0
$t\bar{t}$	440 ± 41	10515 ± 692	2601 ± 196
W + jets (light)	25 ± 6	262 ± 40	5276 ± 417
W + jets (c, b)	11 ± 4	740 ± 148	1595 ± 414
single-top + DY + VV	8 ± 3	734 ± 175	595 ± 153
Total Background	484 ± 44	12252 ± 773	10068 ± 705
Data	463	12157	10204

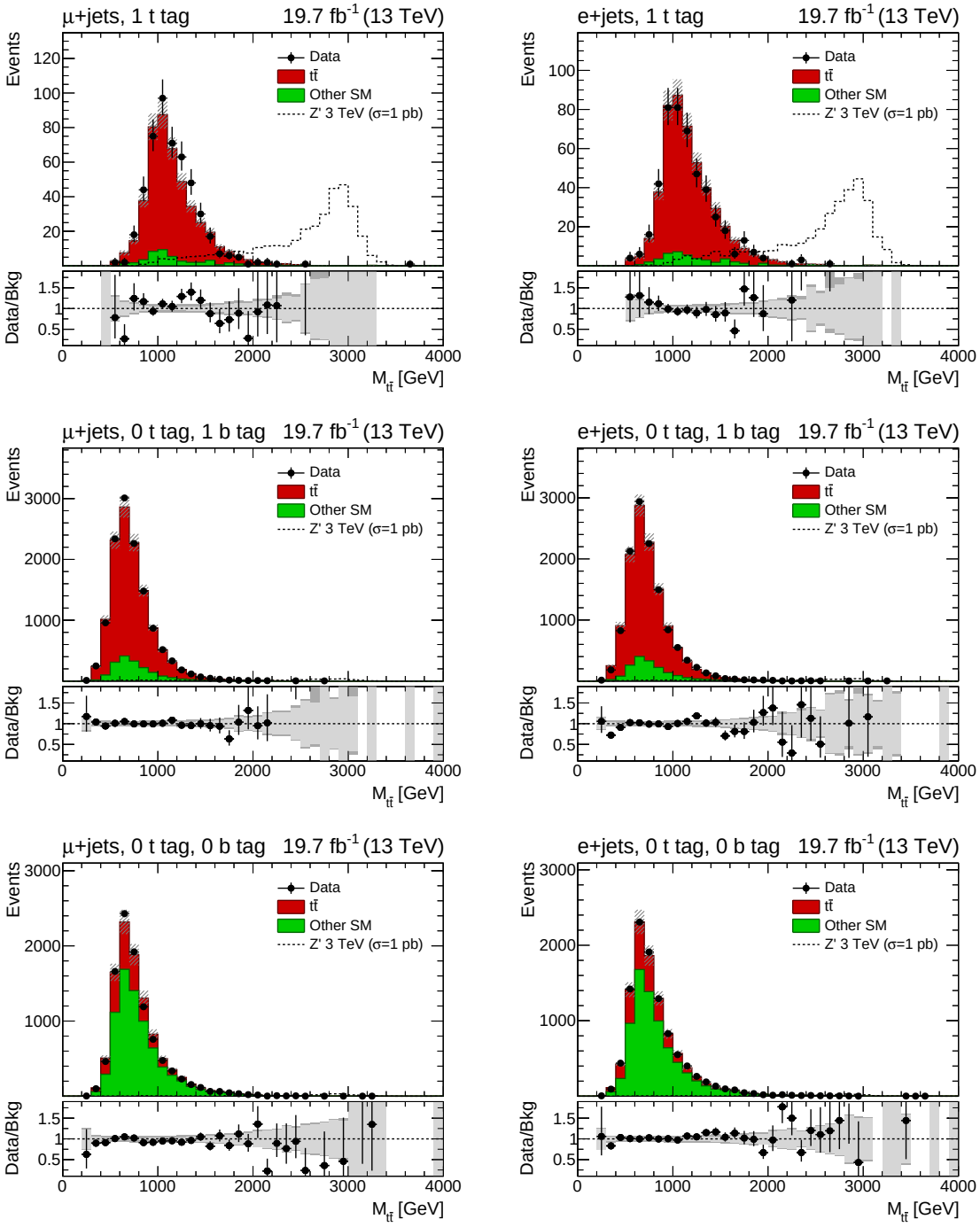


Figure 4.40: Comparison of the observed and expected distributions for the invariant mass of the reconstructed $t\bar{t}$ pair in the six SR categories of the ℓ + jets analysis: (left) μ + jets channel, (right) e + jets channel. From top to bottom, the `T1`, `T0B1`, `T0B0` categories are shown. The background distributions in each plot are obtained after the background-only fit described in the text; the light-grey error band associated to the background prediction includes only the MC statistical uncertainty, while the dark-grey error band also contains all the post-fit systematic uncertainties. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 670.

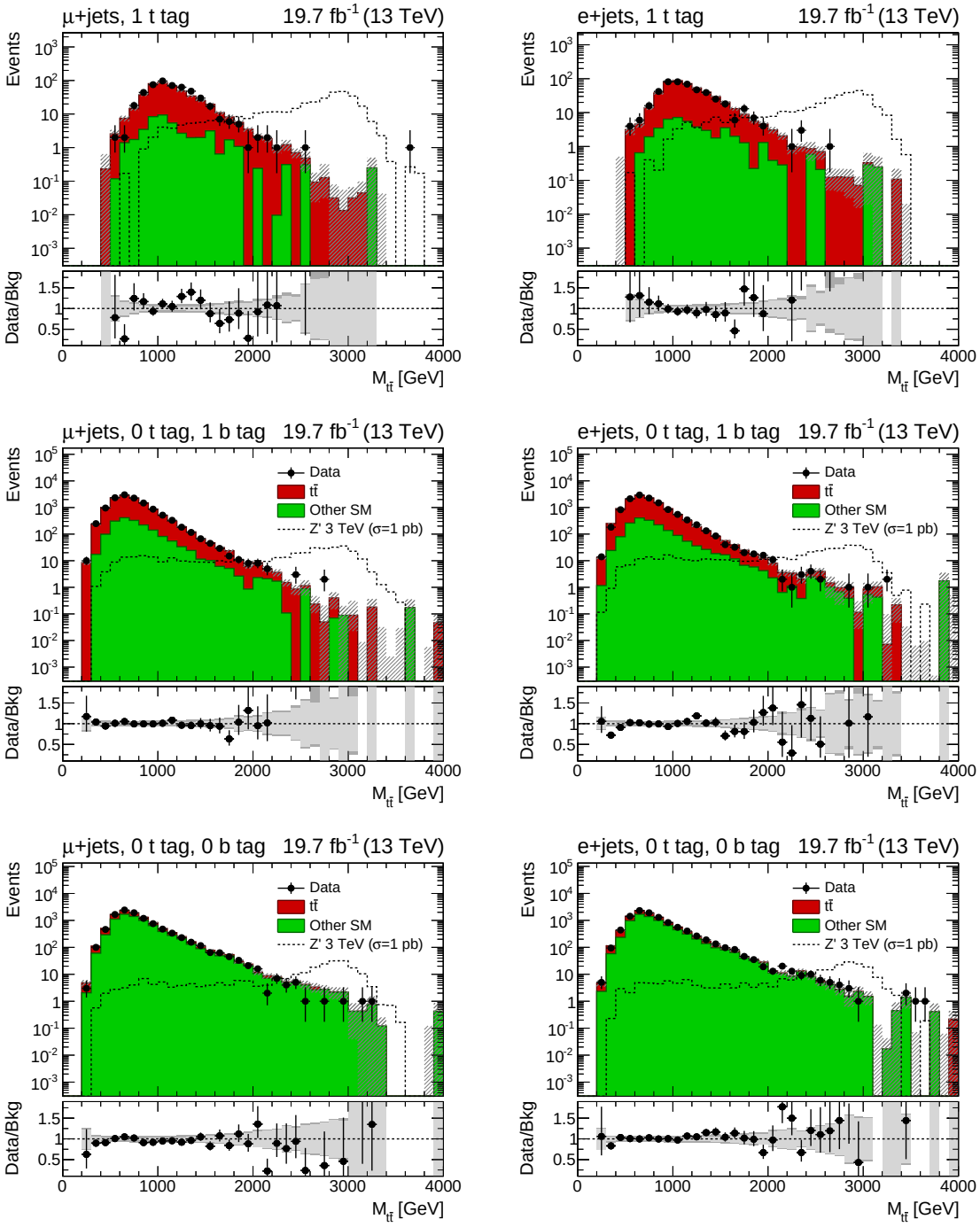


Figure 4.41: Same distributions of Figure 4.40, shown in logarithmic scale.

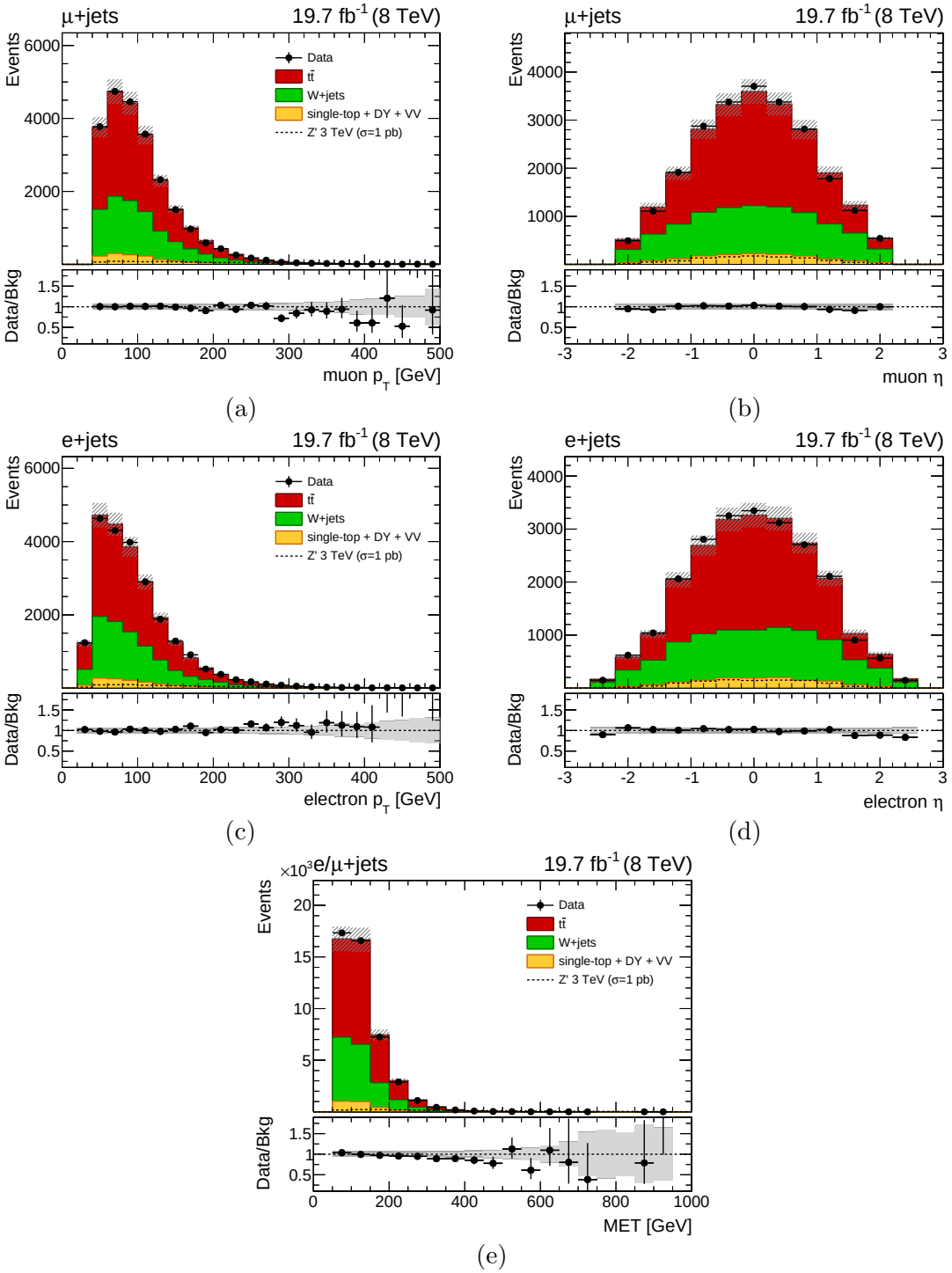


Figure 4.42: Data/MC comparison for the kinematics of the lepton and $\not{E}_T$ in the μ + jets and e + jets SR samples: (a, b) muon p_T and η in the μ + jets channel, (c, d) electron p_T and η in the e+jets channel, (e) $\not{E}_T$ in ℓ+jets events. The expected background is determined after the fit described in the text. The background error band includes the MC statistical error and the overall post-fit systematic uncertainty; in the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 670.

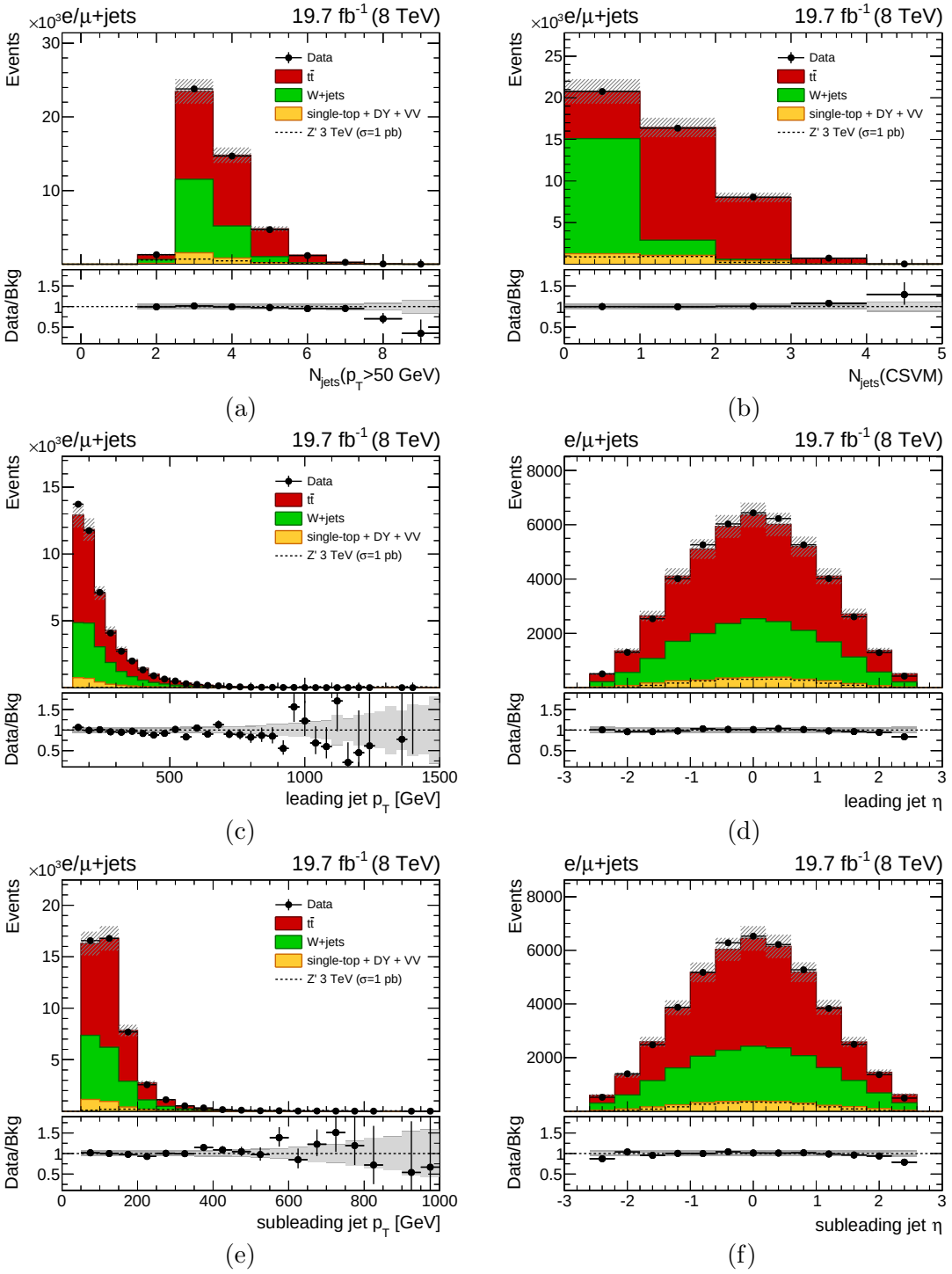


Figure 4.43: Data/MC comparison for the kinematics of AK5 jets in the ℓ + jets SR: (a) number of AK5 jets with $p_T > 50$ GeV, (b) number of AK5 jets passing the CSVM operating point, (c, d) p_T and η of the leading jet, (d, e) p_T and η of the subleading jet. The expected background is determined after the fit described in the text. The background error band includes the MC statistical error and the overall post-fit systematic uncertainty; in the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 670.

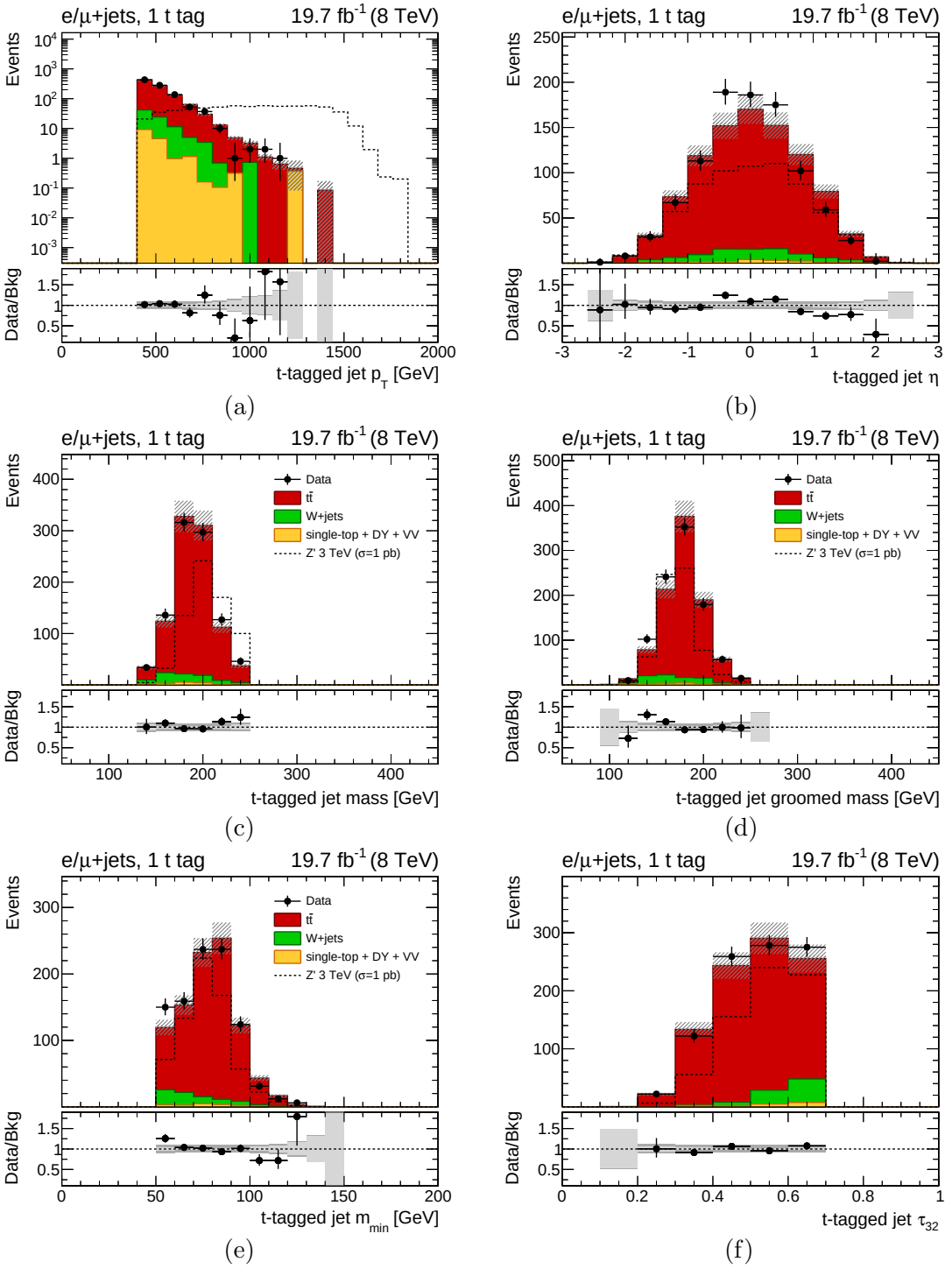


Figure 4.44: Data/MC comparison for the kinematics of the CA8 t-tagged jet in the `T1` category of the ℓ+jets SR: (a) jet p_T, (b) jet η, (c) jet ungroomed mass, (d) jet groomed mass, (e) subjets' minimum pairwise mass, (f) jet τ_{32}. The expected background is determined after the fit described in the text. The background error band includes the MC statistical error and the overall post-fit systematic uncertainty; in the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 670.

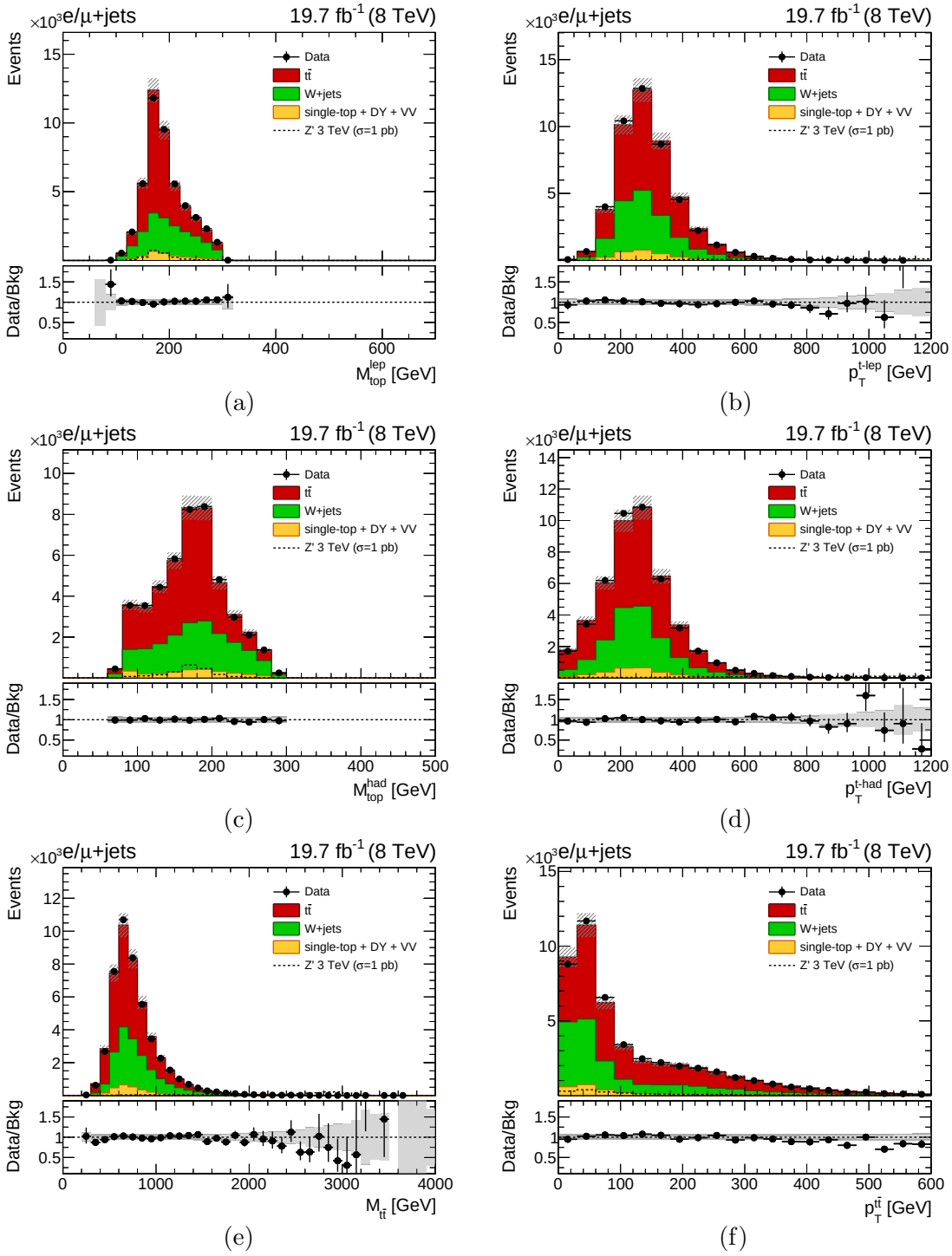


Figure 4.45: Data/MC comparison for the kinematics of the reconstructed $t\bar{t}$ pair in the ℓ+jets SR: (a, b) mass and p_T of the leptonic top quark candidate, (c, d) mass and p_T of the hadronic top quark candidate, (e, f) mass and p_T of the $t\bar{t}$ system. The expected background is determined after the fit described in the text. The background error band includes the MC statistical error and the overall post-fit systematic uncertainty; in the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 670.

4.8 Statistical analysis and results

Since no excess with respect to the background expectation is observed in the data, we proceed to set upper limits on the production cross section of a $t\bar{t}$ resonance for each of the three BSM models considered in the analysis. Limits are set on the signal effective cross section, which corresponds to product of the total production cross section of the resonance, $\sigma(pp \to X)$, times its branching ratio to a $t\bar{t}$ pair, $\mathrm{BR}(X \to t\bar{t})$. Exclusion limits for each signal hypothesis are calculated for a Confidence Level (CL) of 95% using a Bayesian statistical method.

The statistical model used for limit-setting is based on a shape analysis of the reconstructed $M_{t\bar{t}}$ spectrum, which is the kinematic quantity with the best discrimination power between the background and signal distributions. Six $M_{t\bar{t}}$ distributions are used in the statistical analysis, one for each of the event categories defined in Section 4.4.3. Splitting the final event sample in properly designed categories improves the sensitivity of the analysis, as these exclusive categories are characterized by different signal-over-background ratios and they constrain specific systematic uncertainties associated to the model. The statistical analysis is performed using the THETA software package [131].

4.8.1 The statistical model

The test statistic used to determine the exclusion limits is constructed as follows. For each bin i of the observables considered, the expected number of events is given by

$$m_i(\mu, \boldsymbol{\nu}) = b_i(\boldsymbol{\nu}) + \mu \cdot s_i(\boldsymbol{\nu}) \tag{4.11}$$

where b_i (s_i) is the expected number of events predicted by the background (signal-only) model, and μ is a bin-independent factor, commonly referred to as *signal strength*, associated to the normalization of the signal template. The signal distributions used in the statistical model are normalized to a cross section of 1 pb, so that μ can be directly interpreted as the effective cross section of the signal, expressed in picobarns. The latter parameter is the parameter of interest in the statistical analysis, as this is the value for which we determine expected and observed exclusion limits. All the other parameters on which the model depends are denoted by the symbol $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{N_\nu})$. These are referred to as *nuisance parameters* and they are usually associated to statistical or systematic uncertainties on the model's predictions. Given a set of statistically independent measurements $\boldsymbol{n} = (n_1, \ldots, n_N)$, which correspond to the number of observed events in data for each of the N bins of the model, the likelihood function can be constructed from Poisson statistics as follows

$$L(\mu, \boldsymbol{\nu}|\,\boldsymbol{n}) = \prod_{i=1}^{N} P(n_i, m_i) = \prod_{i=1}^{N} \frac{m_i(\mu, \boldsymbol{\nu})^{n_i} \cdot e^{-m_i(\mu, \boldsymbol{\nu})}}{n_i!} \tag{4.12}$$

where the product runs over all the N bins in the model and P corresponds to the probability of observing n_i events in the i-th bin, which is given by the Poisson distribution. Systematic uncertainties are treated as nuisance parameters. Each uncertainty is introduced in the model by multiplying the Poisson likelihood by a prior probability distribution π for the corresponding nuisance parameter. The full likelihood function $\hat{L}$ is thus given by

$$\hat{L}(\mu, \boldsymbol{\nu}|\,\boldsymbol{n}) = L(\mu, \boldsymbol{\nu}|\,\boldsymbol{n}) \prod_{j=1}^{N_\nu} \pi_j(\nu_j) = \prod_{i=1}^{N} \frac{m_i(\mu, \boldsymbol{\nu})^{n_i} \cdot e^{-m_i}}{n_i!} \prod_{j=1}^{N_\nu} \pi_j(\nu_j) \tag{4.13}$$

where π_j corresponds to the prior distribution associated to the j-th nuisance parameter. The prior distribution for the signal strength μ (not included above) corresponds to unity for $\mu \geqslant 0$ and to zero elsewhere, as the signal strength is constrained to be non-negative. Since no prior knowledge of the t-tagging efficiency SF is available for this analysis, the nuisance parameter associated to this systematic is left unconstrained in the statistical analysis, using a flat distribution as its prior. The nuisance parameters associated to systematic uncertainties affecting only the normalization of a given process are assigned a log-normal prior, while the nuisance parameters that affect the shape of the templates in the statistical analysis are assigned a gaussian prior with mean equal to zero and width equal to unity. For each systematic uncertainty affecting the shape of the $M_{t\bar{t}}$ distributions, the $\pm 1\sigma$ input templates correspond to $\nu = \pm 1$ and the distributions associated to other values of the nuisance parameter are extrapolated using the template morphing procedure implemented in THETA.

The uncertainty due to the finite statistics of the MC samples is taken into account using a simplified version of the method proposed by Barlow and Beeston [132]. In the Barlow-Beeston method, one additional nuisance parameter is introduced in the model for every bin and for every MC sample, in order to account for the statistical error of the latter in that given bin. To avoid numerical instabilities in the evaluation of the likelihood due to an exceedingly high number of nuisance parameters, only one additional nuisance parameter is introduced for each bin, with a prior given by a gaussian distribution with mean equal to zero and width equal to the statistical uncertainty on the sum of all the MC samples in that bin. This method is often referred to as the *Barlow-Beeston "light"* method [133]. In order to ensure the reliability of the latter approximation and limit statistical fluctuations in the MC predictions, the $M_{t\bar{t}}$ templates used as input to the shape-based analysis are rebinned such that the statistical error of the total background distribution in each bin is below 30%.

The test statistic used to set limits on the parameter μ corresponds to the Bayesian posterior probability

$$p(\mu|\,\boldsymbol{n}) = \frac{\int d\boldsymbol{\nu}\, \hat{L}(\mu, \boldsymbol{\nu}|\,\boldsymbol{n})}{\int d\mu\, d\boldsymbol{\nu}\, \hat{L}(\mu, \boldsymbol{\nu}|\,\boldsymbol{n})}\,. \tag{4.14}$$

The integral in the numerator of Equation (4.14) is evaluated numerically using a Markov chain Monte Carlo method; the factor in the denominator is only introduced to ensure that the value of $p(\mu|\,\boldsymbol{n})$ lies within zero and unity. As a result, the 95% CL upper limit on the signal strength corresponds to the value $\hat{\mu}$ which satisfies

$$\int_0^{\hat{\mu}} p(\mu|\,\boldsymbol{n})\, d\mu = 0.95\,. \tag{4.15}$$

Observed limits on the signal cross sections are derived in the framework described above by using the $M_{t\bar{t}}$ distributions measured in data. The expected limits and their uncertainties are calculated by performing a large number of pseudo-experiments on toy-data distributions. The median of the distribution of the expected upper limit in these pseudo-experiments and its corresponding 68% (95%) CL interval define, respectively, the expected upper limit and its $\pm 1\sigma$ ($\pm 2\sigma$) error band.

4.8.2 Expected and observed limits on $\sigma(pp \to X \to t\bar{t})$

Exclusion limits are set for three benchmark BSM scenarios, given by a Z' boson with a relative decay width ($\Gamma_{\mathrm{Z}'}/M_{\mathrm{Z}'}$) of 1% and 10%, and a Kaluza-Klein gluon resonance in the RS model. For the latter case, interference effects between the signal and the SM $t\bar{t}$ background are not taken into account here.

Figure 4.46 shows the expected limits on $\sigma \times$ BR for the μ + jets and e + jets channels and their combination. The absolute values of the cross section limits vary depending on the width of the resonance. The discriminating power of the $M_{t\bar{t}}$ distributions between background and signal is higher in the case of a narrow-width $t\bar{t}$ resonance, so tighter cross section limits are obtained for this type of signal. Higher upper limits on $\sigma \times$ BR are found for a Z′ boson with larger decay width and a KK gluon resonance, which has a width ranging between 15% and 20% relative to its mass. A similar sensitivity is found in the μ + jets and e + jets analyses and, as expected, the best exclusion limits are obtained by combining the two channels. The expected cross section limits improve, on average, by roughly 30% when using the ℓ + jets combination, as opposed to using the μ + jets or e + jets channels alone.

The expected and observed cross section limits at 95% CL for the production of a BSM resonance decaying to a $t\bar{t}$ pair are shown in Figures 4.47–4.49 for the three benchmark models and the different analysis channels; the numerical values of the these upper limits are reported in Tables 4.6–4.8. The observed limits are found to be in good agreement with the expected limits within their uncertainties. The largest deviation between the two is found in the mass region between 1.2 TeV and 1.4 TeV, where the observed limits shows an upward fluctuation compared to the expected values. This deviation appears in slightly different measures in both the μ + jets and e + jets channels and it is thus slightly larger in the combined limit. It originates from the data/MC discrepancies observed in the $M_{t\bar{t}}$ distributions in the `T1` category of the μ+jets analysis and in the `T0B1` category of the e+jets analysis. As discussed in Section 4.7, these deviations are not observed consistently in the six ℓ + jets SR categories (and across the two lepton channels for the same category), and they are compatible with the background-only model within 2σ; similarly, the fluctuation in the observed limits is in agreement with the expected limits within 2σ.

We compare the above 95% CL exclusion limits to the theoretical cross sections for a leptophobic Z′ in the topcolor model and a KK gluon in the RS model and we derive exclusion limits on the mass of a $t\bar{t}$ resonance for each of these BSM scenarios. Based on the observed (expected) cross section limits, the ℓ + jets analysis excludes a narrow Z′ resonance with a mass between 0.5 GeV and 2.31 TeV (2.23 TeV), a wide Z′ resonance with a mass between 0.5 GeV and 2.76 TeV (2.65 TeV) and a KK gluon resonance with a mass between 0.7 GeV and 2.52 TeV (2.50 TeV). The complete set of mass limits obtained for the two lepton channels and their combination are reported in Table 4.5.

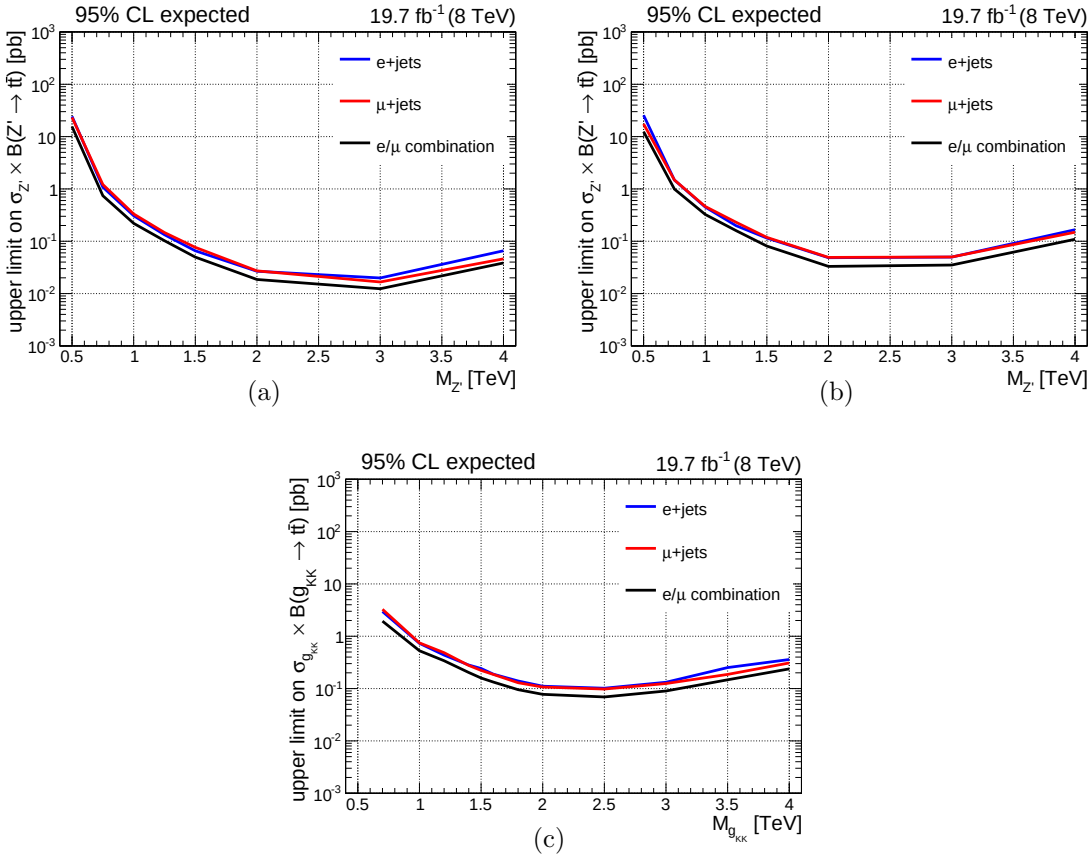


Figure 4.46: Comparison of the 95% CL expected upper limits on $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$ for the μ+jets channel, the e+jets channel and their combination: (a) Z′ boson with 1% width, (b) Z′ boson with 10% width, (c) KK gluon resonance in the RS model.

Table 4.5: Observed and expected excluded mass ranges for the BSM models considered in the statistical analysis. Mass limits are shown for the μ+jets and e+jets channels separately and for their combination.

Excluded mass range [TeV]

signal model	μ + jets channel observed (expected)	e + jets channel observed (expected)	e/μ combination observed (expected)
Z′ (1% width)	0.50 – 2.15 (0.50 – 2.10)	0.50 – 2.16 (0.51 – 2.10)	0.50 – 2.31 (0.50 – 2.23)
Z′ (10% width)	0.50 – 2.61 (0.50 – 2.53)	0.50 – 2.58 (0.50 – 2.53)	0.50 – 2.76 (0.50 – 2.65)
KK gluon	0.70 – 2.42 (0.70 – 2.34)	0.70 – 2.37 (0.70 – 2.33)	0.70 – 2.52 (0.70 – 2.50)

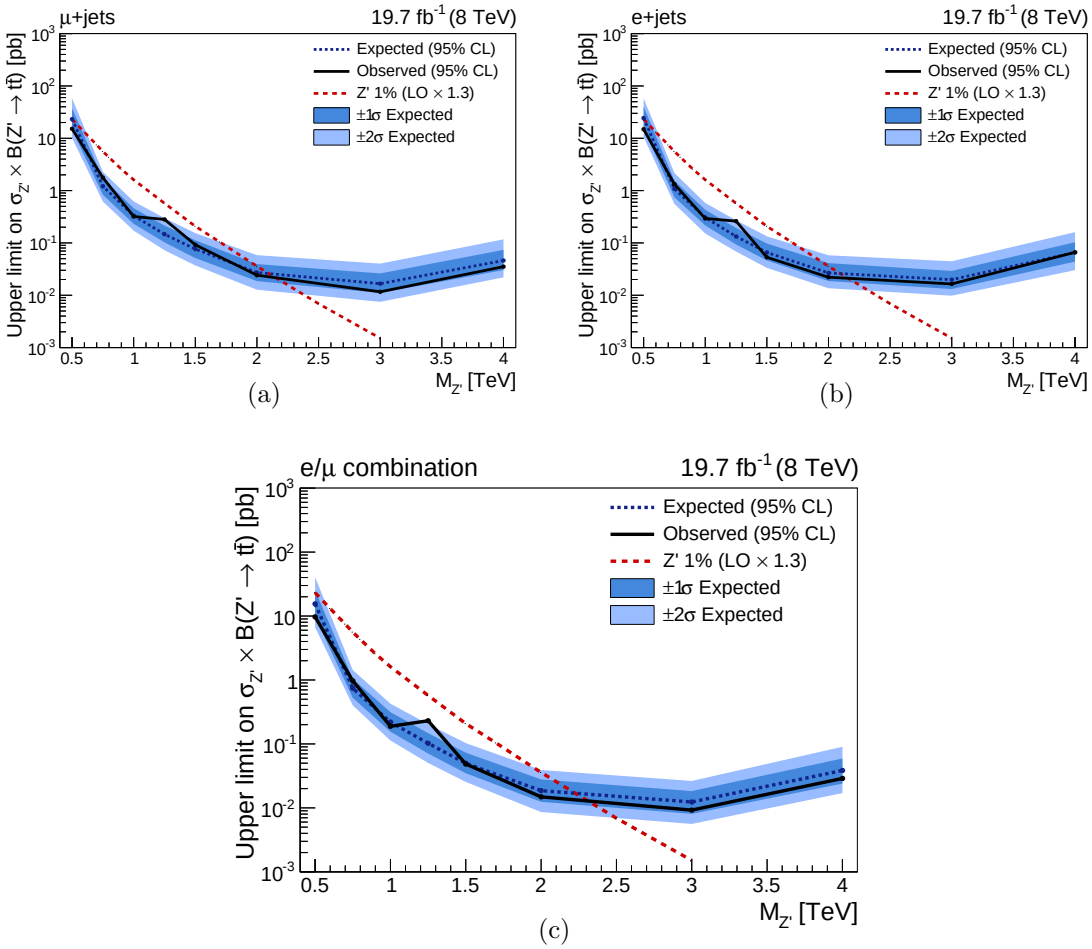


Figure 4.47: 95% CL Bayesian upper limits on the production cross section times branching ratio for a Z′ boson with $\Gamma_{Z'}/M_{Z'} = 1\%$ decaying to $t\bar{t}$, shown as a function of the signal mass hypothesis. The signal cross sections correspond to the LO cross sections for a leptophobic Z′ with a relative width of 1.2% in the topcolor model [112], multiplied by a factor $K = 1.3$ in order to account for higher-order corrections [113]. Limits are shown for the μ+jets (a) and e+jets (b) channels separately and for their statistical combination (c).

Table 4.6: Numerical values for the expected and observed limits shown in Figure 4.47 for the production cross section times branching ratio of a Z′ boson with 1% width decaying to $t\bar{t}$. Limits are given for each of the two analysis channels and their combination.

ℓ + jets **combination**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	15.5	25.2 – 9.7	40.6 – 6.7	9.7
0.75	0.74	1.05 – 0.53	1.43 – 0.40	0.97
1.0	0.22	0.31 – 0.16	0.43 – 0.11	0.19
1.25	0.102	0.148 – 0.071	0.204 – 0.050	0.230
1.5	0.050	0.073 – 0.035	0.103 – 0.025	0.048
2.0	0.0186	0.0278 – 0.0124	0.0391 – 0.0086	0.0149
3.0	0.0124	0.0183 – 0.0081	0.0264 – 0.0056	0.0092
4.0	0.038	0.059 – 0.024	0.090 – 0.017	0.029

μ + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	23.2	37.2 – 14.6	57.9 – 10.1	15.2
0.75	1.21	1.72 – 0.83	2.28 – 0.62	1.77
1.0	0.33	0.46 – 0.23	0.62 – 0.17	0.32
1.25	0.147	0.215 – 0.102	0.299 – 0.072	0.281
1.5	0.076	0.113 – 0.051	0.154 – 0.037	0.092
2.0	0.027	0.039 – 0.019	0.058 – 0.013	0.024
3.0	0.0167	0.0261 – 0.0111	0.0402 – 0.0075	0.0116
4.0	0.046	0.074 – 0.030	0.117 – 0.022	0.035

e + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	24.4	40.3 – 15.2	58.1 – 10.6	14.9
0.75	1.08	1.55 – 0.75	2.17 – 0.56	1.33
1.0	0.31	0.43 – 0.22	0.58 – 0.15	0.29
1.25	0.133	0.193 – 0.094	0.271 – 0.068	0.261
1.5	0.066	0.097 – 0.046	0.134 – 0.033	0.053
2.0	0.027	0.041 – 0.019	0.058 – 0.014	0.022
3.0	0.020	0.029 – 0.013	0.045 – 0.010	0.016
4.0	0.066	0.103 – 0.044	0.161 – 0.030	0.066

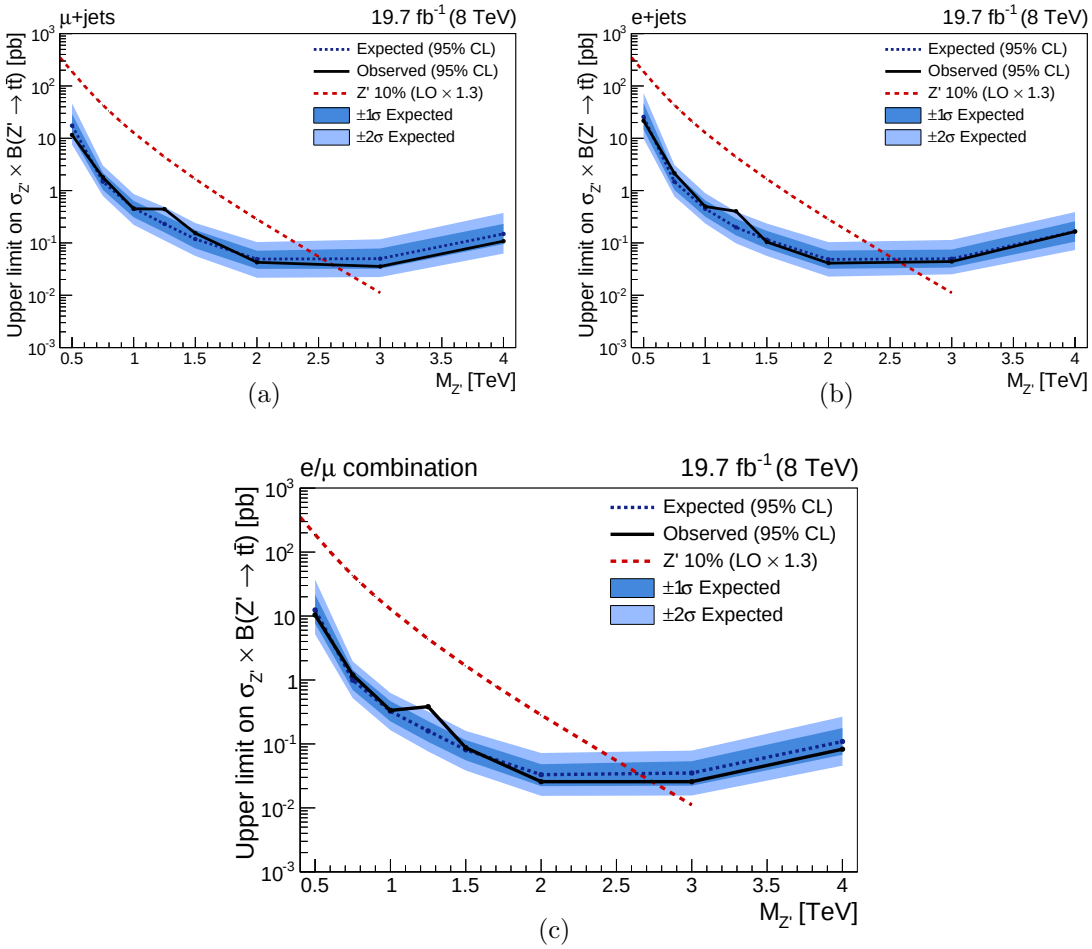


Figure 4.48: 95% CL Bayesian upper limits on the production cross section times branching ratio for a Z′ boson with $\Gamma_{Z'}/M_{Z'} = 10\%$ decaying to $t\bar{t}$, shown as a function of the signal mass hypothesis. The signal cross sections correspond to the LO cross sections for a leptophobic Z′ in the topcolor model [112], multiplied by a factor $K = 1.3$ in order to account for higher-order corrections [113]. Limits are shown for the μ + jets (a) and e + jets (b) channels separately and for their statistical combination (c).

Table 4.7: Numerical values for the expected and observed limits shown in Figure 4.48 for the production cross section times branching ratio of a Z' boson with 10% width decaying to $t\bar{t}$. Limits are given for each of the two analysis channels and their combination.

ℓ + jets **combination**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	12.4	21.8 – 7.6	37.3 – 5.2	10.4
0.75	0.99	1.44 – 0.70	1.95 – 0.52	1.20
1.0	0.33	0.47 – 0.23	0.63 – 0.16	0.33
1.25	0.160	0.227 – 0.108	0.321 – 0.076	0.383
1.5	0.081	0.115 – 0.055	0.160 – 0.038	0.087
2.0	0.033	0.048 – 0.022	0.072 – 0.015	0.026
3.0	0.035	0.054 – 0.022	0.079 – 0.016	0.026
4.0	0.109	0.177 – 0.068	0.266 – 0.046	0.083

μ + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	17.5	28.9 – 11.0	46.5 – 7.6	11.6
0.75	1.48	2.20 – 1.03	3.02 – 0.79	1.80
1.0	0.46	0.64 – 0.31	0.86 – 0.22	0.45
1.25	0.23	0.33 – 0.16	0.49 – 0.11	0.44
1.5	0.119	0.176 – 0.079	0.240 – 0.056	0.153
2.0	0.049	0.071 – 0.032	0.104 – 0.022	0.043
3.0	0.050	0.078 – 0.032	0.117 – 0.022	0.036
4.0	0.149	0.231 – 0.095	0.375 – 0.063	0.108

e + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	25.6	46.9 – 14.6	73.2 – 9.8	21.6
0.75	1.48	2.18 – 1.03	3.10 – 0.77	2.14
1.0	0.45	0.65 – 0.32	0.89 – 0.24	0.50
1.25	0.20	0.29 – 0.14	0.39 – 0.10	0.40
1.5	0.115	0.170 – 0.078	0.235 – 0.056	0.104
2.0	0.048	0.072 – 0.032	0.104 – 0.023	0.041
3.0	0.050	0.075 – 0.034	0.114 – 0.025	0.044
4.0	0.166	0.259 – 0.106	0.389 – 0.073	0.165

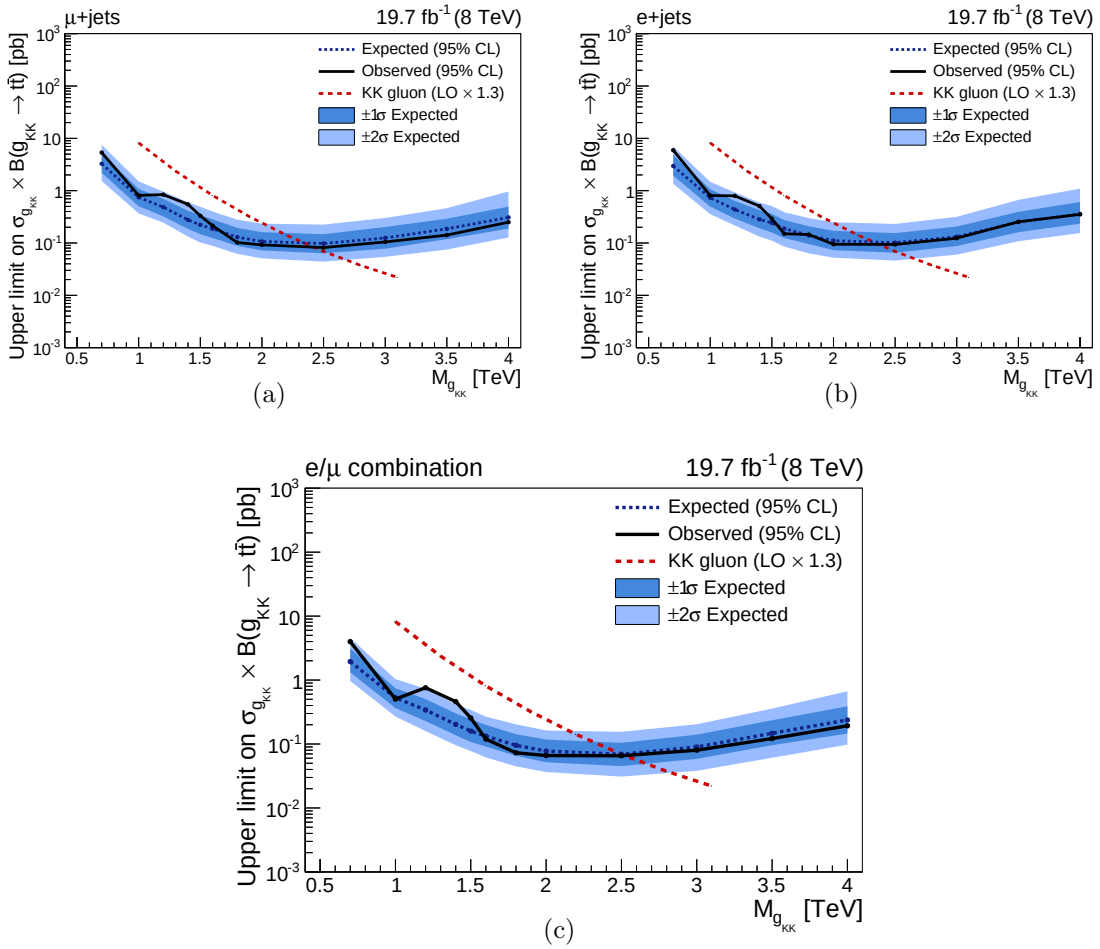


Figure 4.49: 95% CL Bayesian upper limits on the production cross section times branching ratio for a KK gluon decaying to $t\bar{t}$, shown as a function of the signal mass hypothesis. The signal cross sections correspond to the LO cross sections for a KK gluon in the RS model [24], multiplied by a factor $K = 1.3$ in order to account for higher-order corrections [113]. Limits are shown for the μ + jets (a) and e + jets (b) channels separately and for their statistical combination (c).

Table 4.8: Numerical values for the expected and observed limits shown in Figure 4.49 for the production cross section times branching ratio of a KK gluon resonance decaying to $t\bar{t}$. Limits are given for each of the two analysis channels and their combination.

ℓ + jets **combination**

$M_{g_{KK}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.7	1.94	3.11 – 1.32	4.53 – 0.97	3.99
1.0	0.53	0.74 – 0.36	1.03 – 0.27	0.50
1.2	0.34	0.50 – 0.23	0.73 – 0.16	0.76
1.4	0.20	0.30 – 0.14	0.42 – 0.10	0.46
1.5	0.159	0.243 – 0.107	0.341 – 0.078	0.255
1.6	0.133	0.199 – 0.091	0.268 – 0.062	0.119
1.8	0.095	0.142 – 0.065	0.203 – 0.045	0.073
2.0	0.078	0.117 – 0.052	0.163 – 0.036	0.066
2.5	0.069	0.104 – 0.045	0.155 – 0.031	0.065
3.0	0.090	0.140 – 0.059	0.204 – 0.038	0.080
3.5	0.147	0.235 – 0.096	0.358 – 0.061	0.122
4.0	0.24	0.39 – 0.15	0.67 – 0.10	0.19

μ + jets **channel**

$M_{g_{KK}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.7	3.3	5.0 – 2.2	7.5 – 1.5	5.4
1.0	0.75	1.08 – 0.51	1.50 – 0.37	0.81
1.2	0.48	0.70 – 0.32	0.96 – 0.25	0.84
1.4	0.28	0.41 – 0.19	0.60 – 0.13	0.55
1.5	0.22	0.34 – 0.15	0.49 – 0.10	0.33
1.6	0.186	0.283 – 0.125	0.399 – 0.088	0.214
1.8	0.129	0.194 – 0.088	0.278 – 0.062	0.102
2.0	0.107	0.161 – 0.072	0.234 – 0.051	0.092
2.5	0.098	0.147 – 0.063	0.225 – 0.044	0.082
3.0	0.125	0.203 – 0.078	0.301 – 0.054	0.105
3.5	0.19	0.29 – 0.12	0.46 – 0.08	0.14
4.0	0.31	0.50 – 0.19	0.96 – 0.13	0.25

e + jets **channel**

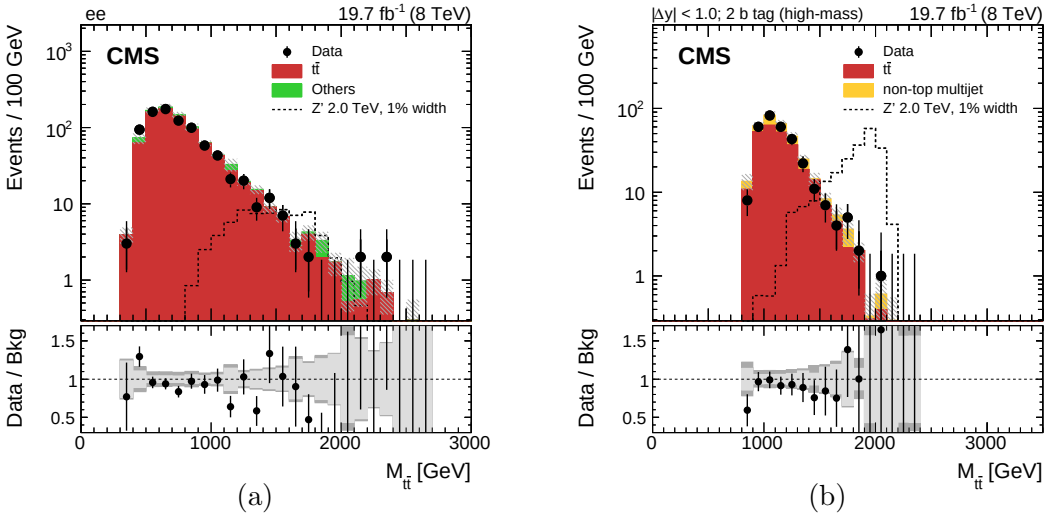
$M_{g_{KK}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.7	3.0	4.8 – 1.9	7.0 – 1.4	5.9
1.0	0.73	1.07 – 0.50	1.47 – 0.35	0.81
1.2	0.44	0.64 – 0.29	0.93 – 0.22	0.80
1.4	0.28	0.43 – 0.19	0.59 – 0.13	0.51
1.5	0.24	0.36 – 0.16	0.51 – 0.11	0.29
1.6	0.19	0.27 – 0.12	0.39 – 0.09	0.15
1.8	0.140	0.209 – 0.096	0.299 – 0.063	0.145
2.0	0.111	0.171 – 0.073	0.249 – 0.052	0.095
2.5	0.102	0.154 – 0.065	0.231 – 0.046	0.094
3.0	0.132	0.207 – 0.087	0.316 – 0.060	0.124
3.5	0.25	0.39 – 0.16	0.67 – 0.11	0.25
4.0	0.36	0.61 – 0.23	1.10 – 0.15	0.35

4.9 Combination of searches for resonant $t\bar{t}$ production at $\sqrt{s} = 8$ TeV

The ℓ + jets analysis described in this chapter has been combined with independent searches for resonant $t\bar{t}$ production at $\sqrt{s} = 8$ TeV in final states with two leptons (dilepton channel) and zero leptons (all-hadronic channel). The results of this combination have been recently published by the CMS Collaboration in Ref. [3]. We refer to the latter publication for a detailed description of each of the analyses included in the combination. This section contains only a summary of the final results.

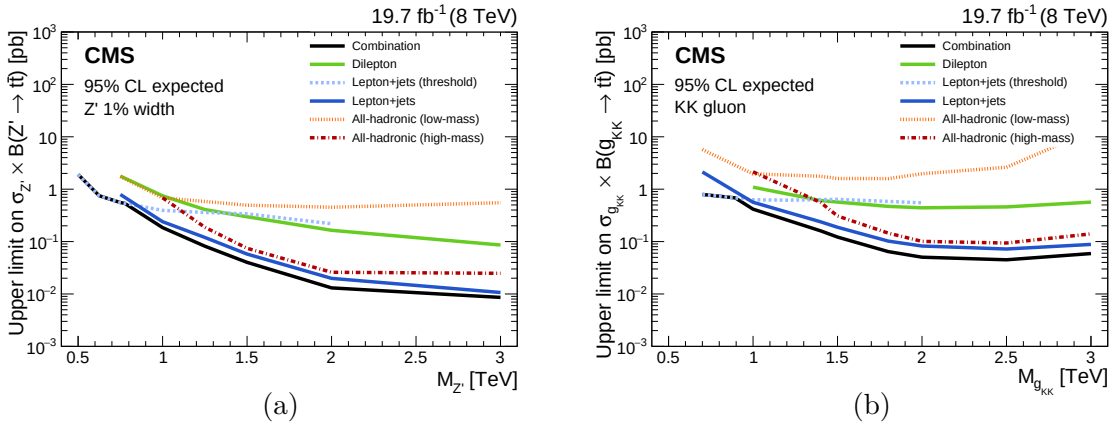
The dilepton analysis selects events with two leptons of opposite electric charge, missing transverse energy and at least one b-tagged jet. The muon and electron selections are optimized for boosted top quark decays using the same identification criteria described in Section 4.3 and 4.4. The analysis considers three independent event samples (ee, $e\mu$ and $\mu\mu$), based on the flavor content of the dilepton system. The all-hadronic analysis includes two independent searches; the first is optimized for high-mass $t\bar{t}$ resonances, while the other focuses on the low-mass region ($M_X \lesssim 1$ TeV). Both analyses select events characterized by a dijet topology, with two large-radius jets both passing a t-tagging selection. The main difference between the two searches comes from the use of two different t-tagging algorithms. The high-mass analysis makes use of CA8 jets with $p_T > 400$ GeV; these jets are required to pass the same t-tagging selection used in the ℓ+jets channel and defined in Section 4.3, based on the CMS Top Tagging algorithm. The low-mass analysis utilizes larger-radius (CA15) jets with $p_T > 200$ GeV passing the HEP Top Tagging algorithm [134] in order to improve the analysis sensitivity for lower-mass $t\bar{t}$ resonances. Finally, the combination also includes an analysis optimized for low-mass $t\bar{t}$ resonances in the ℓ + jets channel [120]; differently from the other searches in the combination, this channel is not statistically independent from the other analyses and it is included in the combination only by superimposing its exclusion limits to those of the other combined searches. Figure 4.50 shows two examples of the $M_{t\bar{t}}$ spectra measured in the dilepton analysis and in the high-mass all-hadronic analysis.

No significant excess in the $M_{t\bar{t}}$ spectra is observed in data, compared to the expected SM background, for any of the searches included in the combination. Upper limits at 95% CL are thus set on the cross section of resonant $t\bar{t}$ production for mass hypotheses between 0.5 TeV and 3 TeV using the invariant mass distribution of the reconstructed $t\bar{t}$ pair. The statistical methods and the signal models considered are the same ones used for the ℓ + jets analysis. Figure 4.51 shows a comparison between the expected limits obtained from each individual channel, together with the expected limits of the full combination. The low-mass ℓ + jets channel provides the best sensitivity for $t\bar{t}$ resonances with masses below 0.75 TeV. For higher values of the signal mass hypothesis, the best expected limits are given by the combination of the all-hadronic, dilepton and boosted ℓ + jets analyses. In the high-mass region, the individual channel with the best expected sensitivity is the ℓ + jets search described in the previous sections; the high-mass all-hadronic analysis reaches a comparable sensitivity for $M_X \gtrsim 2$ TeV. The dilepton and low-mass all-hadronic channels contribute to improve the final sensitivity in the mass region $0.75 \lesssim M_X \lesssim 1.5$ TeV. The combined 95% CL upper limits on $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$ for the three BSM models considered in the analysis are shown in Figure 4.52. Table 4.9 shows the corresponding lower limits on the signal mass hypothesis, derived by comparing the cross section limits to the theoretical predictions for the signal cross sections. Observed and expected limits are found to be in agreement within their respective uncertainties for all the mass hypotheses considered. These represent the best exclusion limits to date on the production cross section of a $t\bar{t}$ resonance at $\sqrt{s} = 8$ TeV.



(a) (b)

Figure 4.50: Reconstructed invariant mass of the $t\bar{t}$ pair for two event categories in the dilepton and all-hadronic analyses: (a) ee events in the dilepton analysis, (b) events in the high-mass all-hadronic analysis with two CA8 t-tagged jets with small separation in rapidity ($|\Delta y_{jj}| < 1.0$) and at least two b-tagged subjets. The uncertainty associated to the background expectation includes both statistical and systematic uncertainties. The data-to-background ratio is shown in the bottom panel of each figure. For the ratio plot, the statistical uncertainty is shown in light gray, while the total uncertainty is shown in dark gray. The expected distribution shown for a Z′ signal with $M_{Z'} = 2$ TeV and $\Gamma_{Z'}/M_{Z'} = 0.01$ is normalized to a cross section of 1 pb, which corresponds to the theoretical cross section multiplied by roughly a factor 30. Both plots are taken from Ref. [3].



(a) (b)

Figure 4.51: Expected 95% CL upper limits on the production cross section times branching fraction for a Z′ boson decaying to $t\bar{t}$ with 1% width (a) and a KK gluon resonance in the RS model (b). The expected limits are shown separately for each sub-channel and the full combination. Both plots are taken from Ref. [3].

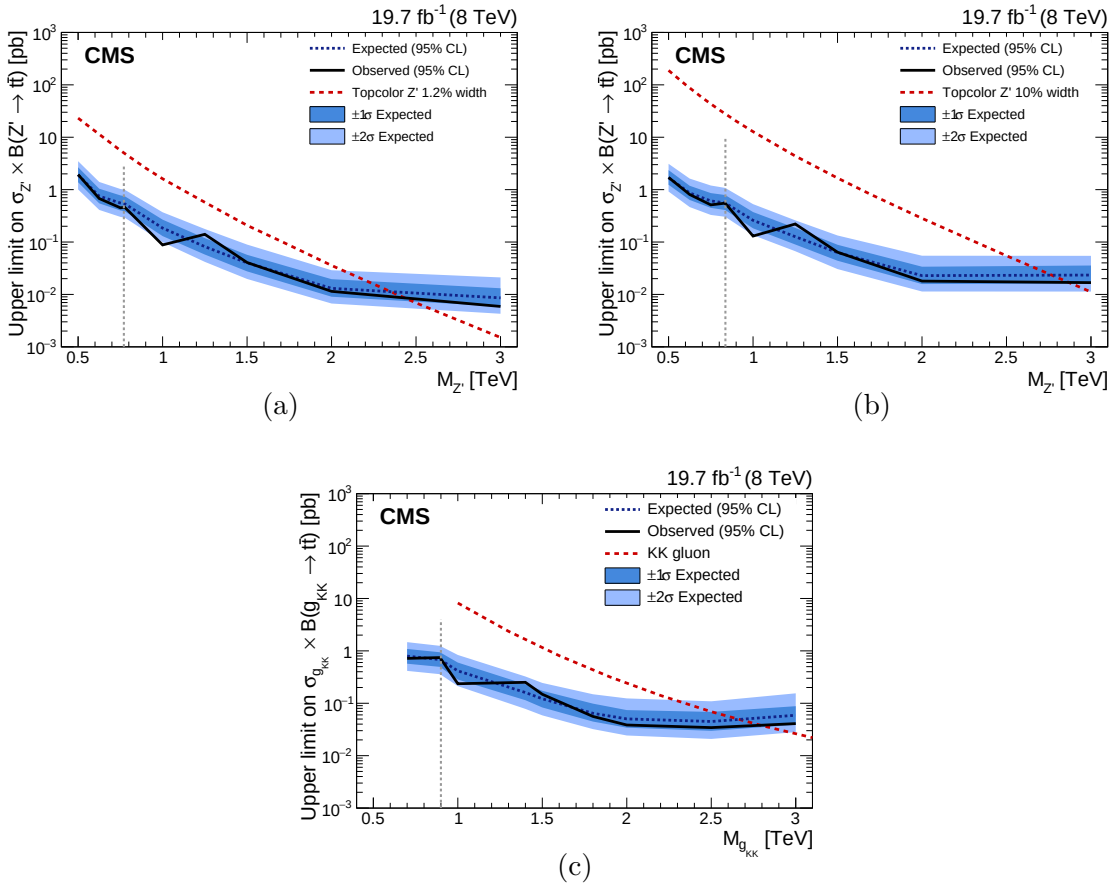


Figure 4.52: Observed and expected 95% CL upper limits on the production cross section times branching fraction to $t\bar{t}$ for a Z′ boson with narrow width (a), with 10% width (b) and a KK gluon resonance in the RS model (c). The vertical dashed line marks the transition between the results of the low-mass ℓ + jets analysis [120] and those from the combination of the other channels. This separation is based on the best expected limit. The limits are shown as a function of the signal mass hypothesis and they are compared to LO calculations for the cross section of a Z′ boson with a relative width of 1.2% and 10% [112] and a KK gluon resonance [24]. These cross sections are multiplied by a factor of 1.3 to account for higher-order corrections [113]. The plots are taken from Ref. [3].

Table 4.9: Observed and expected excluded mass range for each of the three benchmark models considered in the $X \to t\bar{t}$ combination. The values are taken from Ref. [3].

	Excluded mass range [TeV]		
	Z′ ($\Gamma_{Z'}/M_{Z'} = 1.2\%$)	Z′ ($\Gamma_{Z'}/M_{Z'} = 10\%$)	KK gluon
expected	0.5 – 2.4	0.5 – 2.8	0.7 – 2.7
observed	0.5 – 2.4	0.5 – 2.9	0.7 – 2.8

Chapter 5

Search for $t\bar{t}$ resonances in the ℓ + jets channel at $\sqrt{s} = 13$ TeV

This chapter describes a search for resonant $t\bar{t}$ production in the lepton+jets channel with data collected with the CMS detector in pp collisions at $\sqrt{s} = 13$ TeV during the first year of the LHC Run-2 [4]. The large increase in the center-of-mass energy of the LHC collisions opens a new window for the potential discovery of new physics phenomena compared to previous searches. In fact, the cross section for the production of new heavy particles decaying to a $t\bar{t}$ pair in various BSM scenarios is expected to be greatly enhanced going from to the c. o. m. energy of $\sqrt{s} = 8$ TeV of the LHC Run-1 to the unprecedented energy of $\sqrt{s} = 13$ TeV reached at the LHC in Run-2.

5.1 Data-taking conditions in 2015

The first pp collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV were established at the LHC in May 2015, marking the start of the data-taking operations for the LHC Run-2. A total integrated luminosity of approximately 4.2 fb^{-1} of pp collisions data was delivered to the ATLAS and CMS experiments between June and November 2015. For the vast majority of this data-taking period, the LHC delivered proton beams with a record bunch spacing of 25 ns, as opposed to the 50 ns interval employed in the LHC Run-1. The peak instantaneous luminosity registered in 2015 was approximately $5 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The CMS experiment ultimately recorded 3.8 fb^{-1} of pp collisions data, for an overall data-taking efficiency of 91%. The total integrated luminosity recorded by CMS during the 2015 proton-proton run is shown in Figure 5.1. Due to some difficulties in the operation of the CMS magnet, only 79% of this data (3.0 fb^{-1}) were acquired with a fully operational magnet ($B = 3.8$ T), whereas the rest of this data was acquired without any magnetic field in the detector.

5.2 Data sets and simulated samples

Data and MC events have been reconstructed with the official offline software of the CMS experiment (CMSSW, release 7.4). The specifics of the data and simulated samples used for this analysis are given below.

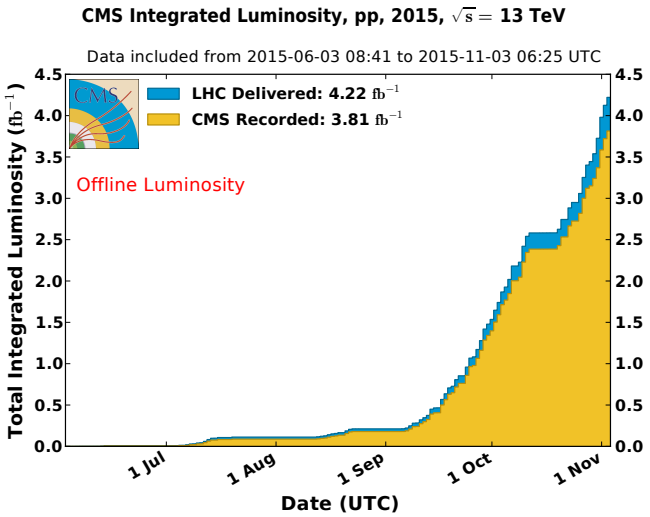


Figure 5.1: Total integrated luminosity delivered to and recorded by the CMS experiment for pp collisions at $\sqrt{s} = 13$ TeV during the course of 2015.

Data sets

The analysis makes use of the data recorded in 2015 by the CMS Collaboration in pp collisions at $\sqrt{s} = 13$ TeV with a bunch spacing of 25 ns and a fully operational magnet at $B = 3.8\,\mathrm{T}$.

The data set considered for this search amounts to a integrated luminosity of 2.6 fb^{-1}. These events correspond to the luminosity sections which pass the offline data-quality certification for all the CMS subdetectors, except for the forward hadron (HF) calorimeter. Inefficiencies in the HF subdetector affected a subsample of 0.3 fb^{-1} of the data recorded by CMS in 2015. This subset of events is retained in the analysis to maximize the size of the input data set. In order to avoid issues related to the suboptimal performance of the HF calorimeter, the missing transverse energy measurement used in the offline analysis is determined without considering the PF candidates reconstructed in the HF calorimeter, as detailed in Section 5.3.1.

Events in the μ + jets analysis are recorded by means of a single-muon trigger, while events in the e + jets channel are triggered using an online selection requiring one electron and at least two hadronic jets. Details on the two triggers strategies are given in Section 5.3.2 and Section 5.4.2.

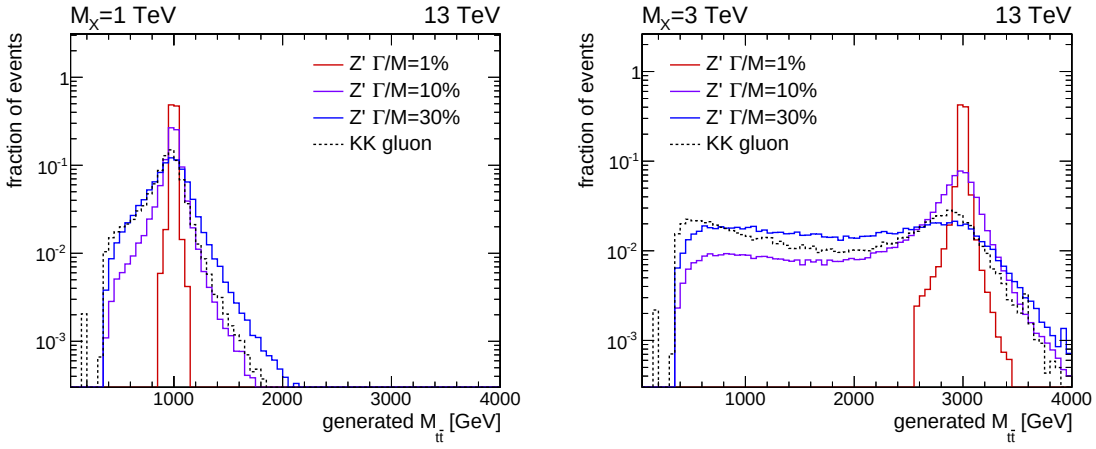
Simulated samples

For all the simulated samples considered in this analysis, events are generated for a c. o. m. energy of 13 TeV and a pp bunch crossing interval of 25 ns; both in-time and out-of-time pileup interactions are included in the event simulation. The PYTHIA-v8.2 generator [135] is used to model the parton showering and hadronization processes; the simulation of the underlying event is performed using the `CUETP8M1` tune [136]. The parton distribution functions are taken from the NNPDF-3.0 PDF parameterization [137]. The simulation of the detector response is based on the GEANT4 package [95].

The production of a Z′ boson decaying to a $t\bar{t}$ pair is simulated at leading-order accuracy with the `MadGraph5_aMC@NLO` package [138]. Signal samples are generated for multiple mass hypotheses, from 0.5 TeV up to 4 TeV. In each of these samples, all the $t\bar{t}$ decay modes are

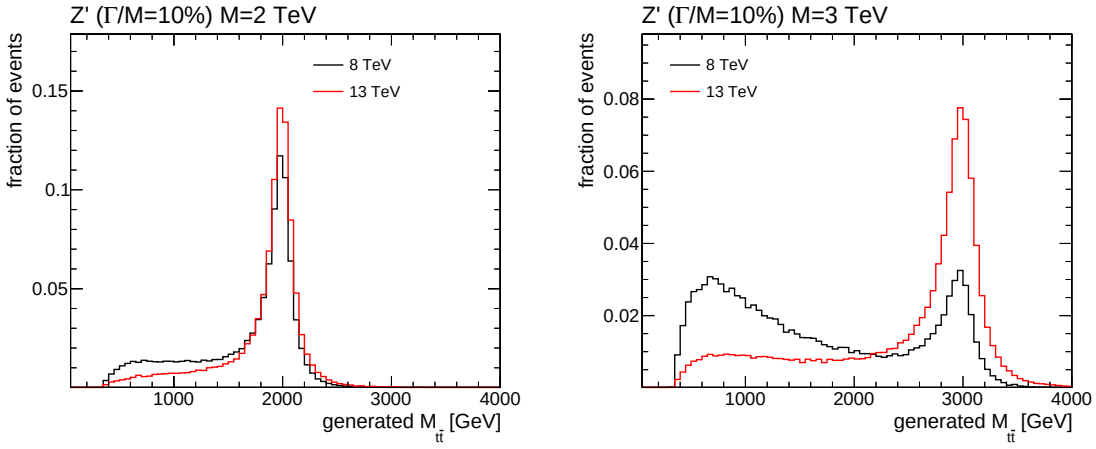
allowed, according to their branching ratios. For this Z′ model, the couplings of the resonance to left-handed and right-handed fermions are set equal to the corresponding couplings of the SM Z boson. The MC matrix element in MadGraph5_aMC@NLO includes QCD corrections for the production of up to three additional partons in the hard-scattering process; the ME-PS matching procedure is performed using the MLM matching scheme [101]. Different width hypotheses are considered for this generic $Z' \to t\bar{t}$ resonance; independent signal samples are produced for three values of the resonance relative width ($\Gamma_{Z'}/M_{Z'}$): 1%, 10% and 30%. The theoretical cross sections for $Z' \to t\bar{t}$ production at the LHC have been calculated up to QCD-NLO accuracy in the topcolor model [139]; these values are used in the final part of this analysis to compare to the cross section limits determined from experimental data. We also test a different BSM scenario given by the production of a Kaluza-Klein excitation of a gluon decaying to a $t\bar{t}$ pair in the RS model. This process is simulated using the PYTHIA-v8.2 event generator for different signal mass hypotheses between 0.5 TeV and 4 TeV. The LO cross sections for these signals are obtained directly from PYTHIA-v8.2 and they are are multiplied by a factor $K = 1.3$ to account for higher-order corrections [113]. Interference effects between the $g_{KK} \to t\bar{t}$ signal and the SM $t\bar{t}$ background are not included in the MC simulation. For all the signal models introduced above, the new particle is a vector boson (spin-1) and it is produced via $q\bar{q}$ annihilation. Figure 5.2 shows the $t\bar{t}$ invariant mass distribution at generator-level, before any selection cut, of the signal models described above for two different mass hypotheses. As already noted in Section 4.2, the generated $M_{t\bar{t}}$ spectra for wide BSM signals (10%- and 30%-width Z′ boson and KK gluon) are characterized by a lower-mass tail coming from off-shell production, due to the convolution of the parton-level cross section with the parton distribution functions; this effect becomes increasingly prominent for higher values of the resonance's width. On the other hand, the enhancement of the parton luminosities at the increased center-of-mass energy of $\sqrt{s} = 13$ TeV leads to a sizeable reduction of off-shell production compared to Run-1; a comparison between the generated $t\bar{t}$ masses at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV can be found in Figure 5.3. The theoretical cross sections for the $Z' \to t\bar{t}$ and $g_{KK} \to t\bar{t}$ signals are shown in Figure 5.4.

The simulation of SM $t\bar{t}$ production is performed at QCD-NLO accuracy using the POWHEG-v2 generator [96, 97, 98]. In addition to an inclusive $t\bar{t}$ sample, dedicated samples binned in the generator-level mass of the $t\bar{t}$ system are utilized to increase the MC statistical power in the high-$M_{t\bar{t}}$ region. The $t\bar{t}$ simulation is normalized to the inclusive cross section calculated at NNLO+NNLL accuracy in QCD using the TOP++2.0 package [99, 140]. Single-top production is also simulated using NLO generators: POWHEG-v1 is used for tW production, while top quark production in the s-channel and t-channel is simulated with MadGraph5_aMC@NLO at NLO. Predictions for the single-top cross section in the s and t channels have been calculated at NLO accuracy in QCD using the HATHOR-v2.1 program [141]; for the tW channel, we use the inclusive cross section calculated up to approximate NNLO in QCD [32, 33]. The production of a V = (W, Z) boson in association with jets is simulated at LO with MadGraph5_aMC@NLO, including QCD corrections for the production of up to four additional partons in the MC matrix element. A combination of exclusive samples is used to improve the statistics for the simulation of these processes; specifically, the V + jets samples used in this analysis are binned in H_T^{GEN}, where H_T^{GEN} denotes the scalar sum of the transverse momenta of the partons in the MC matrix element. The inclusive cross section for W + jets and Z + jets, which are used to normalize these samples, are calculated at NNLO in QCD with FEWZ-v3.1 [102]. Simulated samples for diboson production (WW, WZ and ZZ) are generated with PYTHIA-v8.2. The cross sections for WZ and ZZ are calculated up to QCD-NLO accuracy using MCFM-v6.6 [103]; for the WW cross section, we use the NNLO



(a) (b)

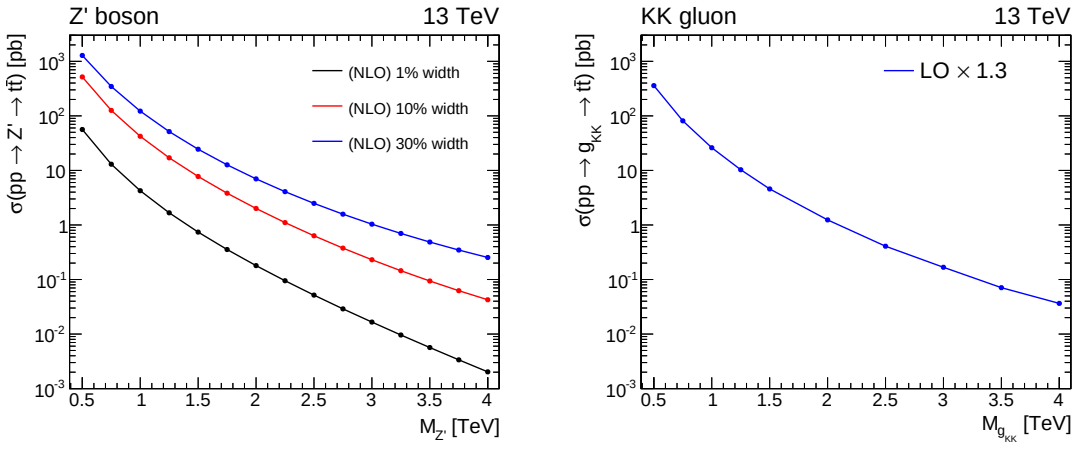
Figure 5.2: Invariant mass of the $t\bar{t}$ system at generator level for the BSM models considered in the analysis, for a mass hypothesis of 1 TeV (a) and 3 TeV (b). Every distribution is normalized to unity.



(a) (b)

Figure 5.3: Comparison of the invariant mass of the generated $t\bar{t}$ pair at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV: (a) $Z' \to t\bar{t}$ signal with $M_{Z'} = 2$ TeV and 200 GeV width, (b) $Z' \to t\bar{t}$ signal with $M_{Z'} = 3$ TeV and 300 GeV width. Every distribution is normalized to unity.

calculation given in Ref. [142]. Simulated samples of QCD-multijet production are generated using PYTHIA-v8.2. We use exclusive QCD samples binned in $\hat{p}_T$, where $\hat{p}_T$ denotes the momentum transfer in the MC matrix element. These samples are used in the analysis only for optimization studies on lepton identification. Table 5.1 provides a summary of the MC samples used in this analysis for the simulation of the SM backgrounds, together with the cross section assigned to each process.



(a) (b)

Figure 5.4: Theoretical cross sections at 13 TeV for the BSM $t\bar{t}$ resonances considered in the analysis. Left: cross sections for a leptophobic $Z' \to t\bar{t}$ signal in the topcolor model calculated up to QCD-NLO accuracy [139]. Right: cross sections for a $g_{KK} \to t\bar{t}$ signal calculated at LO with PYTHIA-v8.2 and multiplied by a factor $K = 1.3$ to account for higher-order corrections [113].

Table 5.1: Simulated samples and cross sections used in the analysis for SM processes.

process	samples	MC generator	cross section [pb]
$t\bar{t}$ + jets	$t\bar{t}$ (inclusive)	POWHEG-v2	831.8
	$t\bar{t}$ (0.7 TeV $< M_{t\bar{t}}^{\text{GEN}} <$ 1 TeV)	POWHEG-v2	76.6
	$t\bar{t}$ ($M_{t\bar{t}}^{\text{GEN}} >$ 1 TeV)	POWHEG-v2	20.6
W + jets	W($\to \ell\nu$) + jets (0.1 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.2 TeV)	MadGraph5_aMC@NLO (LO)	1627.5
	W($\to \ell\nu$) + jets (0.2 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.4 TeV)	MadGraph5_aMC@NLO (LO)	435.2
	W($\to \ell\nu$) + jets (0.4 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.6 TeV)	MadGraph5_aMC@NLO (LO)	59.2
	W($\to \ell\nu$) + jets (0.6 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.8 TeV)	MadGraph5_aMC@NLO (LO)	14.6
	W($\to \ell\nu$) + jets (0.8 TeV $< H_{\text{T}}^{\text{GEN}} <$ 1.2 TeV)	MadGraph5_aMC@NLO (LO)	6.7
	W($\to \ell\nu$) + jets (1.2 TeV $< H_{\text{T}}^{\text{GEN}} <$ 2.5 TeV)	MadGraph5_aMC@NLO (LO)	1.6
	W($\to \ell\nu$) + jets (2.5 TeV $< H_{\text{T}}^{\text{GEN}}$)	MadGraph5_aMC@NLO (LO)	0.04
Z + jets	Z($\to \ell\ell$) + jets ($M_{\ell\ell}^{\text{GEN}} >$ 50 GeV, 0.1 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.2 TeV)	MadGraph5_aMC@NLO (LO)	181.3
	Z($\to \ell\ell$) + jets ($M_{\ell\ell}^{\text{GEN}} >$ 50 GeV, 0.2 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.4 TeV)	MadGraph5_aMC@NLO (LO)	50.4
	Z($\to \ell\ell$) + jets ($M_{\ell\ell}^{\text{GEN}} >$ 50 GeV, 0.4 TeV $< H_{\text{T}}^{\text{GEN}} <$ 0.6 TeV)	MadGraph5_aMC@NLO (LO)	7.0
	Z($\to \ell\ell$) + jets ($M_{\ell\ell}^{\text{GEN}} >$ 50 GeV, 0.6 TeV $< H_{\text{T}}^{\text{GEN}}$)	MadGraph5_aMC@NLO (LO)	2.7
single top	s-channel ($t \to b\text{W} \to b\ell\nu$)	MadGraph5_aMC@NLO (NLO)	3.4
	t-channel ($t \to b\text{W} \to b\ell\nu$)	MadGraph5_aMC@NLO (NLO)	70.7
	tW-channel	POWHEG-v1	71.2
diboson	WW	PYTHIA-v8.2	118.7
	WZ	PYTHIA-v8.2	47.1
	ZZ	PYTHIA-v8.2	16.5

5.3 Analysis strategy

5.3.1 Object reconstruction

The offline event reconstruction is based on the CMS Particle Flow algorithm [66], described in Section 3.1. This sections details the specific methods used to define the different physics objects considered in the analysis.

5.3.1.1 Primary vertices and pileup correction in simulated events

Primary vertices are determined based on the kinematics of the reconstructed tracks in the event, using the same methodology described in Section 4.3.1.

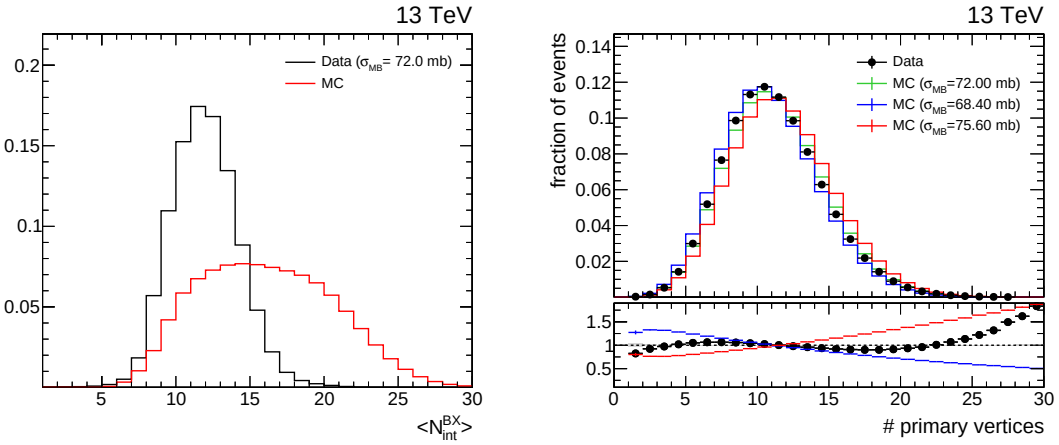
These vertices are ranked based on their $\sum p_T^2$ value, but instead of considering the standard sum over all tracks associated to a given vertex, a slightly more sophisticated method is used. Tracks not identified as muons or electrons are clustered into jets and the full collection of tracks is used to estimate the charged-only $\not{E}_T$ in the event, which quantifies the energy contribution from neutral candidates. The $\sum p_T^2$ score of each primary vertex is calculated using tracks associated to muons and electrons and higher-level objects such as track-jets and charged-only $\not{E}_T$. This vertex sorting increases the efficiency of selecting the correct primary vertex, especially in events where a large portion of the transverse energy comes from neutral particles or high-p_T quarks, which are typically reconstructed as a collection of low-p_T tracks. The vertex with the highest $\sum p_T^2$ value is identified as the primary vertex in the event.

Simulated events are reweighted to match the average number of pileup interactions measured in the 2015 data-taking campaign. The mean number of interactions per bunch crossing ($N_{\text{int}}^{\text{BX}}$) in data events is estimated by multiplying the average instantaneous luminosity per bunch crossing by the total inelastic, or minimum-bias, cross section for pp collisions at $\sqrt{s} = 13$ TeV. This quantity in data corresponds to the average number of interactions generated in simulated events. The ratio of the two distributions is used to derive data/MC corrections as a function of $N_{\text{int}}^{\text{BX}}$, which are used to reweight the simulation. This is shown in Figure 5.5(a). A minimum-bias cross section of 72 mb is used to derive these correction factors [143]. Residual differences between data and reweighted MC events in the distribution of primary vertices are covered by considering a 5% variation on the value chosen for the minimum-bias cross section. This uncertainty is later propagated to the final results of the analysis as a systematic error. Figure 5.5(b) shows the number of primary vertices measured in data compared to three MC distributions, reweighted using different values of the total inelastic cross section.

5.3.1.2 Muons

The muon candidates in the analysis are required to have $p_T^\mu > 50$ GeV and $|\eta^\mu| < 2.1$. The choice of these thresholds is driven by the requirements in the online trigger selection, which is discussed in Section 5.4.2. Each muon candidate is also required to satisfy the following set of identification cuts:

- the muon is associated to a PF candidate;
- the muon is either a global-muon or a tracker-muon;
- a lower cut on the fraction of valid tracker hits;



(a) (b)

Figure 5.5: (a) Average number of interactions per bunch crossing simulated in MC events and measured in data, assuming a minimum-bias cross section of 72 mb. Each distribution is normalized to unity. (b) Number of reconstructed primary vertices, before any event selection, comparing data and simulated events, after correcting the pileup distribution in MC using three different values for the minimum-bias cross section. Each distribution is normalized to unity. The lower-hand plot shows the ratio of each of the above distributions with respect to the central MC prediction (green line).

- at least one of the following two requirements is met:

  1. the muon global track satisfies all the following conditions:
     - the muon candidate is a global-muon;
     - an upper cut on the normalized χ^2 of the muon global track;
     - an upper cut on the χ^2 of the position match between standalone-muon and tracker-muon;
     - an upper cut on the χ^2 from the kink-finder on the inner track;
     - a loose lower cut on the muon segments' compatibility;
  2. a tight lower cut on the muon segments' compatibility.

This set of requirements is referred to in the following as the muon Medium-ID [76]. Compared to the Tight-ID method used in the Run-1 version of this analysis, the Medium-ID leads to an higher efficiency for prompt muons and a comparable rejection rate for non-prompt ones. A detailed comparison of the performance of these two identification methods is given in Section 5.4.1.

5.3.1.3 Electrons

The electron candidates in the analysis are required to have $p_T^e > 50$ GeV and $|\eta_{\rm SC}^e| < 2.5$, where $\eta_{\rm SC}^e$ stands for the pseudorapidity of the electron supercluster. The cut on pseudorapidity is dictated by the geometrical acceptance of the electromagnetic calorimeter, whereas the choice of the offline p_T threshold is driven by the trigger selection used in the $e+$jets channel, as detailed in Section 5.4.2.

Each electron candidate is required to pass a set of tight identification cuts. This electron ID selection does not include any requirement related to the electron isolation. The rationale of this cut-based ID selection is very similar to the analogous method developed in CMS for Run-1 [77], but the cut thresholds have been re-optimized for the different data-taking conditions expected in Run-2.

The electron cut-based ID is defined by the following set of cuts:

- upper cut on a measure of the spread in η of the electron supercluster;
- upper cut on the absolute difference in η between the electron track and the electron supercluster;
- upper cut on the absolute difference in ϕ between the electron track and the electron supercluster;
- upper cut on the ratio of the hadronic energy associated to the electron and the energy of the electron's supercluster measured in the electromagnetic calorimeter;
- upper cut on a measure of the difference ($|1/E - 1/p|$) between the electron's ECAL energy and its momentum;
- upper cut on the impact parameter of the electron GSF track with respect to the interaction vertex;
- upper cut on the distance along the beam direction (z-axis) between the electron GSF track and the interaction vertex;
- upper cut on the number of expected missing hits in the inner tracker;
- veto on electrons coming from photon conversions.

The performance of this identification requirements is studied in simulated events in Section 5.4.1, which also includes a comparison to other electron identification methods.

5.3.1.4 Small-radius and large-radius hadronic jets

Two jet collections are employed in the analysis. Jets are reconstructed with the anti-k_{T} (AK) algorithm for two different choices of the cone parameter: $R = 0.4$ and $R = 0.8$. We refer to the first (second) set of jets as AK4 (AK8) or small-radius (large-radius) jets. Both collections are built using the Charged Hadron Subtraction method, in which the input to the clustering sequence corresponds to the list of all PF candidates not marked as pileup-hadrons.

Jet energy corrections (JEC) are applied to each jet candidate, in both data and simulation, to correct for neutral pileup contributions, η-dependence and p_T-dependence of the detector response [114, 115]. An additional set of JECs is applied only in data events to account for residual differences between data and simulation. Different JECs are used for AK4 and AK8 jets. The jet energy resolution (JER) in MC events is also corrected, in order to match the slightly worse energy resolution measured in data using dijet and γ + jets events [114, 116].

The removal of leptons from the two jet collections reconstructed in the event (jet-lepton cleaning) is performed as follows. In the case of AK4 jets, jet-lepton cleaning is performed by matching the PF-references of the jet constituents and those of the PF candidates associated

to each lepton: when a matching is found, the four-momentum of the PF candidate associated to the lepton is subtracted from the uncorrected four-momentum of the corresponding jet; the resulting 4-vector is finally rescaled with the appropriate JECs. The candidate-based approach used in this search represents an improvement with respect to the geometric matching used for the cleaning of small-radius jets in the Run-1 version of this analysis. As already described in Section 4.3, this type of cleaning is necessary for the correct reconstruction of small-radius jets, because in the leptonic decay of an high-p_T top quark a prompt lepton is expected to have a very small angular separation from a small-radius jet coming from the decay of a b-quark; the conventional requirement of a minimum angular separation between leptons and small-radius jets to perform jet-lepton cleaning would thus be not suited for this search. For the AK8 collection, jet-lepton cleaning is implemented by simply requiring each AK8 jet to have a minimal angular separation ($\Delta R > 0.8$) from every lepton candidate. In the case of AK8 jets, this simpler procedure is inherently equivalent to the one based on PF-references in the context of this analysis, because hadronic decays of boosted objects, like W bosons or top quarks, are not expected to overlap with prompt leptons in $t\bar{t}$ events.

AK4 jet candidates are required to have $p_T > 30$ GeV and $|\eta| < 2.4$. Large-radius (AK8) jet candidates, which are used for the reconstruction of hadronically decaying high-p_T top quarks, are required to have $p_T > 500$ GeV and $|\eta| < 2.4$. Jet candidates in both collections are also required to pass minimal jet quality criteria (PFJet-ID Loose working point) [114].

Jet b-tagging

Small-radius jets are used to identify bottom quark decays. A new version of the CSV algorithm (CSV-v2) is used [86]. The algorithm assigns to each jet a single discriminator, proportional to the probability that the jet originated from the decay of a b-quark. In this analysis an AK4 jet is said to be b-tagged if it passes the medium working point of the CSV-v2 tagger.

Jet t-tagging

A jet t-tagging algorithm is used in the analysis in order to identify AK8 jets originating from the hadronic decay of a top quark. This method allows to identify high-p_T top quarks whose decay products are highly collimated and that are reconstructed in the detector as a single large-radius jet with substructure properties. The t-tagging selection applied in this analysis has been developed by the CMS Collaboration [88]. It is based on a combination of cuts on the following two variables:

- Jet softdrop mass, M_{SD}: the mass of the four-vector sum of the jet constituents selected by the softdrop algorithm [90], with $z_{cut} = 0.1$ and $\beta = 0$ (see Section 3.5.4). The application of softdrop declustering leads to the removal of soft and wide-angle particles associated to the jet. As a result, the softdrop mass, compared to the plain jet mass, provides greater discrimination power between light- and t-flavor jets, and better mass resolution for the latter jets. The t-tagging selection includes the following cut on the jet softdrop mass: 110 GeV $< M_{\mathrm{SD}} <$ 210 GeV.

- N-subjettiness ratio $\tau_{32} \equiv \tau_3/\tau_2$: the N-subjettiness variable τ_N, defined in Equation (3.9), is a substructure variable designed to quantify to compatibility of a reconstructed jet with the decay of N partons [91]. Jets coming top quarks are expected to have low values of τ_3 and relatively higher values for other hypotheses ($N \neq 3$). The

τ_{32} ratio is the variable giving the best discrimination between t-jets and light-flavor jets. AK8 jets passing the t-tagging selection are required to have $\tau_{32} < 0.69$.

The performance of this t-tagging algorithm is studied in Section 5.4.3.

5.3.1.5 Missing transverse energy

The measurement of the transverse energy imbalance in the event is performed with a PF-based method. The measured missing E_T is corrected in order to properly take into account the effect of the JECs in the MET measurement (Type-1 MET). Since part of the dataset considered for this analysis was affected by problems in the Hadron-Forward (HF) calorimeter, we use the MET measured without taking into consideration the PF candidates reconstructed in this subdetector. In simulated events, corrections to the energy resolution of jets are also propagated to the $\not{E}_T$ measurement.

5.3.2 Event preselection

The signature of interest is given by the production of a boosted $t\bar{t}$ pair decaying in the semileptonic channel, where one W boson from a top quark decays leptonically into a lepton (electron or muon) and a neutrino, and the other W boson decays hadronically. Selected events are thus required to have one high-p_T lepton, two or more high-p_T hadronic jets and missing transverse energy. The inclusive selection on the hadronic jets in the event, to be compared with the conventional requirement of at least four jets for a resolved $t\bar{t} \to \ell$ + jets final state, allows to cover event topologies characterized by the production of boosted top quarks that decay hadronically and that are reconstructed in the detector by only one or two jets.

Events in data affected by different types of detector-related problems are rejected with specific data-quality requirements. These are used to filter out events affected by anomalous noise in the hadronic calorimeters and ECAL crystals, events affected by beam halo interactions and events without any well-reconstructed primary vertex. These filters are also applied to simulated events and their overall impact is found to be very small, as they reject approximately 0.3% of all MC events.

The specific selection cuts used in the μ + jets and e + jets channels are listed in Section 5.3.2.1 and Section 5.3.2.2, respectively. The μ + jets selection is essentially identical to the one employed in the Run-1 version of the analysis. On the other hand, the e + jets selection is characterized by a slightly different set of requirements, compared to both the μ + jets channel in Run-2 and the e + jets selection in Run-1. In the muon channel, an unprescaled single-muon trigger with an online p_T-threshold of 45 GeV is used, similarly to the strategy adopted in Run-1. In the e + jets channel, we require already at the HLT level the presence of one electron with $p_T > 45$ GeV and two jets with, respectively, $p_T > 200$ GeV and $p_T > 50$ GeV. Compared to the corresponding Run-1 selection, higher p_T thresholds on the HLT jets were necessary to keep this trigger unprescaled during data-taking. As a consequence, higher p_T-thresholds on the two leading jets are applied in the e + jets analysis to ensure an optimal trigger efficiency, leading to a much tighter event selection compared to the μ + jets analysis. A tighter cut is also applied on $\not{E}_T$ in the e + jets channel ($\not{E}_T > 120$ GeV), compared to the μ + jets channel ($\not{E}_T > 50$ GeV). This is done to reduce to a negligible amount the non-$t\bar{t}$ background coming from QCD-multijet production. As already discussed in Chapter 4, this reducible background is harder to suppress in the e + jets rather than in the μ + jets final state, because of the higher probability of having a jet mis-identified

as an electron at the reconstruction level. The triangular-cuts selection used in Run-1 to complement the $\not\!\!E_T$ requirement in the $e+$jets channel is not used in this analysis: this is due to the fact that the discrimination power of this cut is reduced when using the $\not\!\!E_T$ measurement without HF candidates, as opposed to using the full $\not\!\!E_T$ like in the Run-1 analysis. Several approaches have been tested to optimally reduce the QCD background, for example by combining requirements on $\not\!\!E_T$, $H_{\mathrm{T}}^{\mathrm{lep}}$, p_T^{W} and triangular-cuts; using only a tighter selection on the missing transverse energy is ultimately found to be the most effective approach. Similarly to the Run-1 analysis, the final discrimination between prompt and non-prompt leptons is implemented via the so-called lepton 2D-cut. While this is the same method already used in Run-1, the cut parameters have been re-optimized for Run-2; these optimization studies, together with comparisons between the lepton 2D-cut and alternative methods, are detailed in Section 5.4.1.

5.3.2.1 Event preselection in the $\mu+$jets channel

The event selection in the $\mu+$jets analysis is based on the following requirements.

- The event passes the unprescaled single-muon HLT path `HLT_Mu45_eta2p1_v*`;
- exactly one muon candidate ($p_T^{\mu} > 50$ GeV, $|\eta^{\mu}| < 2.1$, Medium ID) is required in the event; events with additional muon or electron candidates are vetoed;
- at least one AK4 jet with $p_T > 150$ GeV and $|\eta| < 2.4$ is required;
- a second AK4 jet with $p_T > 50$ GeV and $|\eta| < 2.4$ is required;
- $\not\!\!E_T > 50$ GeV and $H_{\mathrm{T}}^{\mathrm{lep}} > 150$ GeV, where $H_{\mathrm{T}}^{\mathrm{lep}} \equiv \not\!\!E_T + p_T^{\mu}$;
- lepton 2D-cut: the event is accepted if either $\Delta R(\mu, j) > 0.4$ or $p_{T,\mathrm{rel}}(\mu, j) > 20$ GeV, where j stands here for the ΔR-nearest jet to the muon candidate. The quantity $p_{T,\mathrm{rel}}(\mu, j)$ is the relative transverse momentum of the leading muon with respect to its closest jet in ΔR. Both quantities are calculated with respect to all AK4 jets with $p_T^j > 15$ GeV and $|\eta^j| < 3$.

5.3.2.2 Event preselection in the $e+$jets channel

The event selection in the $e+$jets analysis is based on the following requirements.

- The event passes the unprescaled single-electron HLT path `HLT_Ele45_CaloIdVT_GsfTrkIdT_PFJet200_PFJet50_v*`;
- exactly one electron candidate ($p_T^e > 50$ GeV, $|\eta_{\mathrm{SC}}^e| < 2.5$, Tight cut-based ID without isolation cut) is required in the event; events with additional muon or electron candidates, are vetoed;
- at least one AK4 jet with $p_T > 250$ GeV and $|\eta| < 2.4$ is required;
- a second AK4 jet with $p_T > 70$ GeV and $|\eta| < 2.4$ is required;
- $\not\!\!E_T > 120$ GeV;
- the electron candidate in the event is required to pass the same lepton 2D-cut defined above for the $\mu+$jets selection.

5.3.3 Kinematic reconstruction of the $t\bar{t}$ system

The $t\bar{t}$ system is reconstructed by assigning the four-vectors of the reconstructed final-state objects (charged lepton, $\not{E}_T$, jets) to the leptonic and hadronic legs of the $t\bar{t}$ decay.

The charged lepton and $\not{E}_T$ are assigned to the leptonically decaying top quark, interpreting the reconstructed $\not{E}_T$ as the transverse component of the neutrino momentum. The only experimental quantity that cannot be reconstructed directly is the longitudinal component of the neutrino momentum, $p_{z,\nu}$. This quantity can be determined indirectly using the following quadratic equation

$$p^{\pm}_{z,\nu} = \frac{\mu p_{z,\ell}}{p^2_{T,\ell}} \pm \sqrt{\frac{\mu^2 p^2_{z,\ell}}{p^4_{T,\ell}} - \frac{E^2_\ell p^2_{T,\nu} - \mu^2}{p^2_{T,\ell}}} \,, \tag{5.1}$$

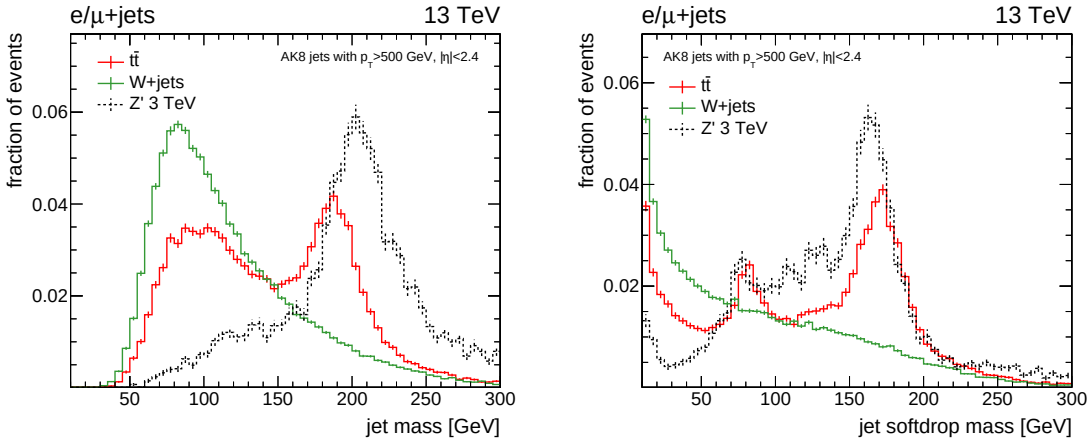
where p_ℓ and p_ν are the four-momenta of the charged lepton and the neutrino, respectively, and $\mu = \frac{M^2_{\ell+\nu}}{2} + p_{T,\ell}\, p_{T,\nu} \cos\Delta\phi$. In order to determine a value for $p^{\pm}_{z,\nu}$, we assign to the mass of the lepton+$\not{E}_T$ system the value of the W boson mass, $M_{\ell+\nu} = M_{\mathrm{W}} = 80.4$ GeV. Equation (5.1) can have either zero, one, or two real solutions. In the absence of a real solution, the real part of the complex solutions is used. If there are two real solutions, both cases are tested, effectively doubling the number of $t\bar{t}$ hypotheses in the event.

The assignment of jets to the two top quarks is determined as follows. For events without a t-tagged AK8 jet, a list of all possible assignments of AK4 jets is constructed, i.e. each jet is assigned to either the leptonically decaying top quark, the hadronically decaying top quark or neither of the two. For events in which one AK8 jet passes the t-tagging selection, the $t\bar{t}$ hypotheses are reconstructed with a simpler methodology: the t-tagged jet is the only object assigned to the hadronically decaying top quark and all AK4 jets with $\Delta R < 1.2$ from the hadronic top quark are not considered further in the kinematic reconstruction. For each hypothesis, the top quark four-momenta are given by the sum of the four-momenta of all the reconstructed objects assigned to it. Hypotheses with no jets assigned to either the leptonic or hadronic top quark are not retained.

A χ^2 discriminator is used to identify the best $t\bar{t}$ hypothesis in the event. The selection criterion is based on the fact that the reconstructed top quark masses are expected to be close to the top quark mass. This is implemented by constructing the following discriminator

$$\chi^2 = \left[\frac{M^{\mathrm{lep}}_{\mathrm{top}} - \bar{m}^{\mathrm{lep}}_{\mathrm{top}}}{\sigma^{\mathrm{lep}}_M}\right]^2 + \left[\frac{M^{\mathrm{had}}_{\mathrm{top}} - \bar{m}^{\mathrm{had}}_{\mathrm{top}}}{\sigma^{\mathrm{had}}_M}\right]^2 \,, \tag{5.2}$$

where $M^{\mathrm{lep}}_{\mathrm{top}}$ corresponds to the mass of the reconstructed leptonic top quark and $M^{\mathrm{had}}_{\mathrm{top}}$ is a mass value associated to the reconstructed hadronic top quark. In events without a t-tagged jet, $M^{\mathrm{had}}_{\mathrm{top}}$ corresponds to the invariant mass of the small-radius jets assigned to the hadronic top quark; in events where a t-tagged jet is present, the value of $M^{\mathrm{had}}_{\mathrm{top}}$ is taken from the groomed (softdrop) mass of the t-tagged jet, as opposed to its plain mass. This is done because the groomed mass, compared to the ungroomed mass, is much less dependent on the jet p_T and it provides greater discrimination power between signal and non-top background events. This is exemplified in Figure 5.6, which shows a comparison between the two types of jet masses for the W + jets and $t\bar{t}$ backgrounds and a $\mathrm{Z}' \to t\bar{t}$ signal. It is worth noting that, in the case of true $t\bar{t}$ events, the softdrop jet mass also shows a second peak around 80 GeV, due to AK8 jets associated to the hadronic decay of the W boson. For each event, we select the hypothesis with the smallest χ^2 value as the best $t\bar{t}$ hypothesis; the other hypotheses are not considered further in the analysis.



(a) (b)

Figure 5.6: Ungroomed (a) and groomed (b) masses of AK8 jet candidates (before any t-tagging requirement) for W + jets, $t\bar{t}$ and a Z′ signal with $M_{Z'} = 3$ TeV in events passing the ℓ + jets preselection. Each distribution is normalized to unity.

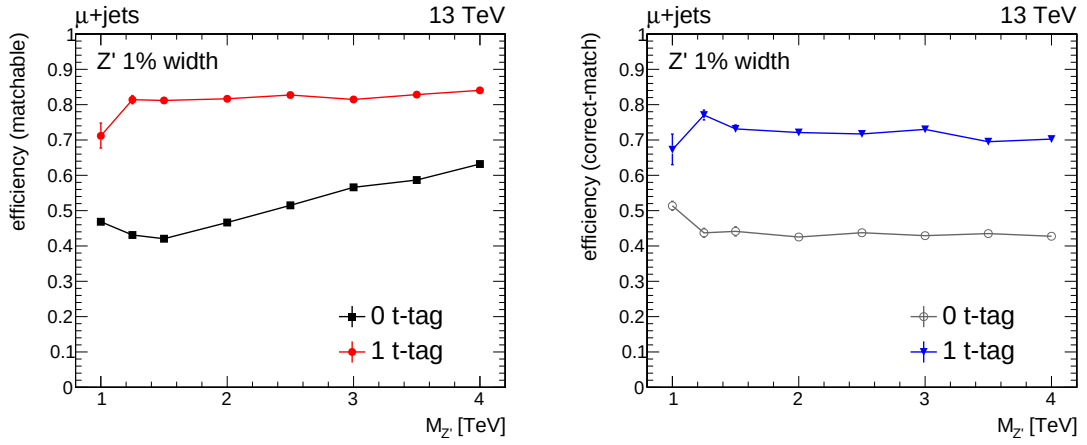
Parameterization of the χ^2 discriminator

The parameters $\bar{m}_{\text{top}}^{\text{lep}}$, σ_M^{lep}, $\bar{m}_{\text{top}}^{\text{had}}$ and σ_M^{had} in the χ^2 discriminator of Equation (5.2) are determined as follows. We consider MC events in which the semileptonic decay of a $t\bar{t}$ is simulated at generator-level. In addition, we require these events to pass the ℓ + jets preselection described in Section 5.3.2 and to satisfy minimal requirements that ensure the event kinematics is compatible to that of a $t\bar{t}$ event. For events without a t-tagged jet and only two (three) AK4 jets, we require one jet to have $m_j > 120$ (50) GeV, i.e. have a jet mass close to the top quark (W boson) mass; no additional requirements are applied to events with a t-tagged jet or with at least four AK4 jets. The MC events satisfying all the above requirements are defined as *reconstructable events*. In these events we reconstruct all the possible $t\bar{t}$ hypotheses and we determine the *correct* $t\bar{t}$ hypothesis using a geometrical matching between the reconstructed objects of the $t\bar{t}$ hypothesis and the decay products of the generated $t\bar{t}$ pair. This matching is implemented as follows.

- **Leptonic top quark**: the AK4 jet associated to the leptonic top quark decay is required to have $\Delta R < 0.4$ with respect to the b-quark from the generated leptonic top quark. The angular separation between reconstructed and generated lepton must be lower than $\Delta R = 0.1$ and the angular distance in the transverse plane between the generated neutrino and the missing transverse energy must be lower than $\Delta\phi = 0.4$.

- **Hadronic top quark**: if the event contains a t-tagged jet, the latter is required to have $\Delta R < 0.8$ with respect to one of the generated top quarks. For events without a t-tagged jet, each AK4 jet associated to the hadronic top quark decay is required to have $\Delta R < 0.4$ with respect to a quark from the generated top quark; each quark has to be matched to one of these jets and more than one quark can be matched to the same jet.

If more than one $t\bar{t}$ hypothesis passes these requirements, the correct $t\bar{t}$ hypothesis is defined as the one with the smallest $\sum \Delta R$ between reconstructed objects and generated particles. We call *matchable events* those for which a correct $t\bar{t}$ hypothesis exists and we define as *matching efficiency* the percentage of reconstructable events that are matchable. We also



(a) (b)

Figure 5.7: Matching efficiency (a) and correct-χ^2 efficiency (b) in the μ + jets channel for the narrow-width Z′ samples, as a function of the signal mass hypothesis. The matching efficiency corresponds to the percentage of reconstructable events for which the correct $t\bar{t}$ hypothesis can be determined. The correct-χ^2 efficiency corresponds to the fraction of matchable events in which the correct $t\bar{t}$ hypothesis coincides with the hypothesis selected by the χ^2 discriminator. The efficiencies are determined separately for events with and without a t-tagged jet ("1 t-tag" and "0 t-tag" events).

define the *correct-χ^2 efficiency* as the fraction of matchable events for which the correct $t\bar{t}$ hypothesis corresponds to the $t\bar{t}$ hypothesis selected by the χ^2 discriminator.

The matching and correct-χ^2 efficiencies for the narrow-width Z′ signals are shown in Figure 5.7. The matching requirements on lepton and neutrino have a very similar impact on the 0 t-tag and 1 t-tag categories: for both categories, the efficiency for the lepton matching cut alone is very high, above 99.5%, while the neutrino-$\not\!\!E_T$ requirement accounts for an efficiency loss ranging between 10% and 15%. The difference in the final matching efficiencies across the two event categories are due to the requirements on the decay products of the hadronic top quark. In events with a t-tagged jet, the matching requirements on jets lead to an additional reduction of less than 10% in the overall efficiency, thanks to the very high-p_T of the top quark decay products. In the 0 t-tag category, the jet-quark matching requirements reduce the final matching efficiency significantly, due to the acceptance limits introduced by the object reconstruction; for example, the matching fails if one quark from the top quark decay is produced with a relatively low transverse momentum and it is not reconstructed as a jet candidate (which has a p_T-threshold of 30 GeV). This acceptance effect becomes less important as the value of the signal mass hypothesis increases. The correct-χ^2 efficiency remains stable as a function of the signal mass hypothesis in both the 0 t-tag and 1 t-tag categories; the higher efficiency obtained for the latter category is due to the simpler topology of the event and the limited number of possible jet assignments.

The M_{lep} and M_{had} distributions for the correct $t\bar{t}$ hypothesis are fitted with a gaussian function to estimate the mean value and width of the masses of the reconstructed hadronic and leptonic top quarks, to be used in the discriminator of Equation (5.2). These distributions are shown in Figures 5.8–5.9 for the μ + jets channel; very similar results are obtained in the e + jets channel. The mean values obtained from the gaussian fits are found to be compatible, within their resolution, across different signal mass hypotheses.

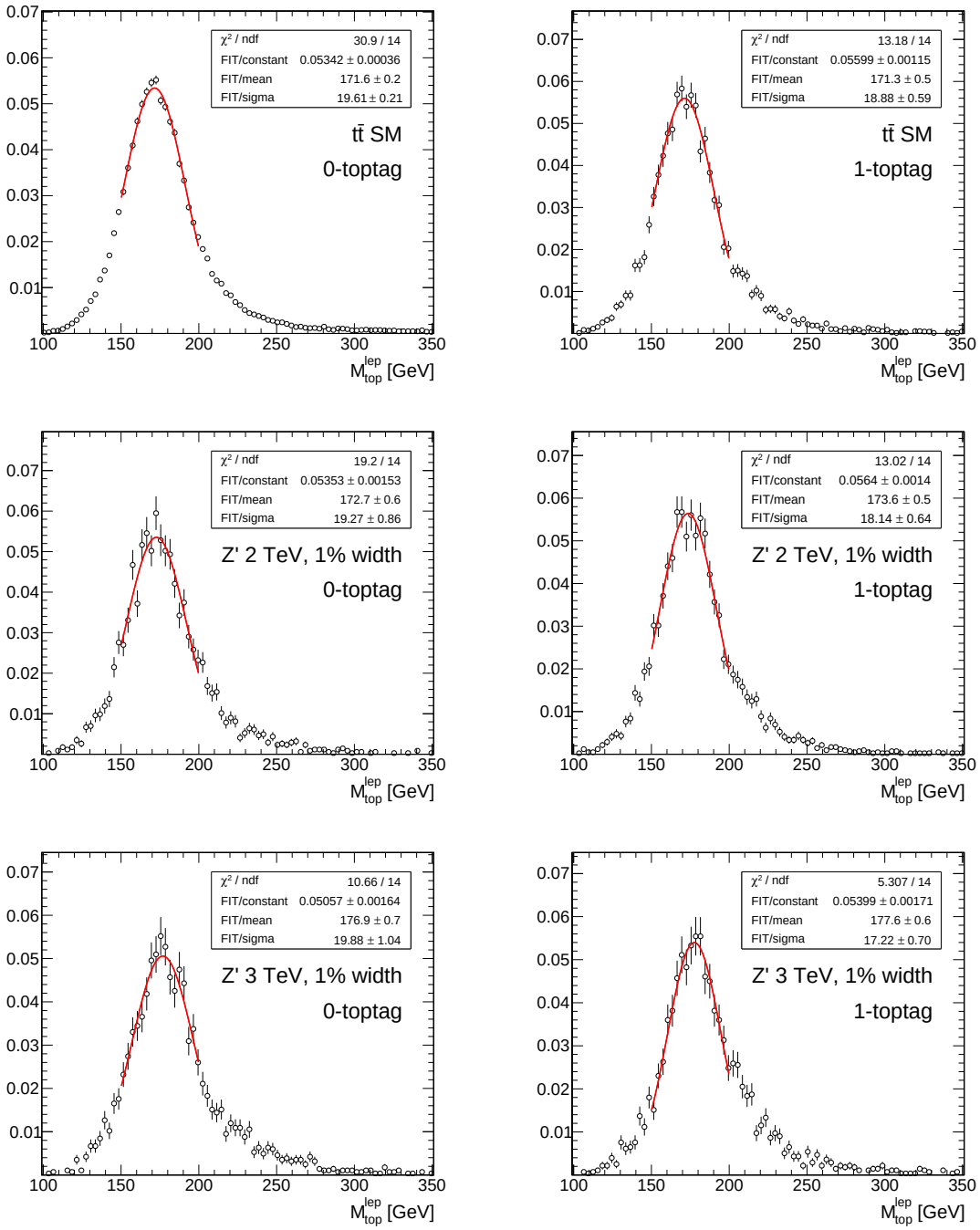


Figure 5.8: Mass of the reconstructed leptonic top quark for the correct $t\bar{t}$ hypothesis (defined by ΔR-matching between generated particles and reconstructed objects) in the $\mu+$jets channel. Events without (with) a t-tagged jet are shown on the left (right). The red line in each plot corresponds to the fit to a gaussian function. Distributions are shown for SM $t\bar{t}$ production (top) and a 1%-width Z′ boson with $M_{Z'} = 2$ TeV (center) and $M_{Z'} = 3$ TeV (bottom).

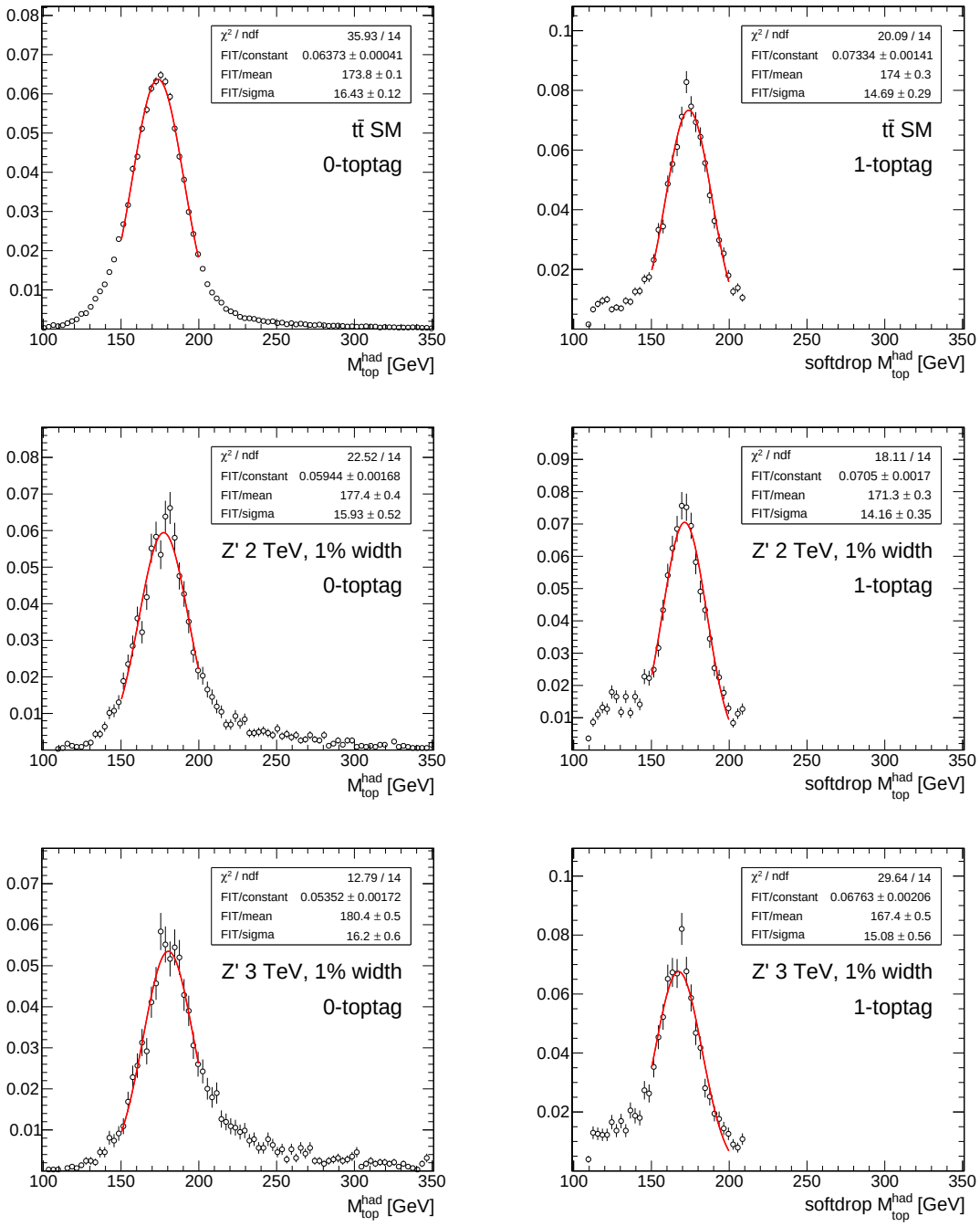
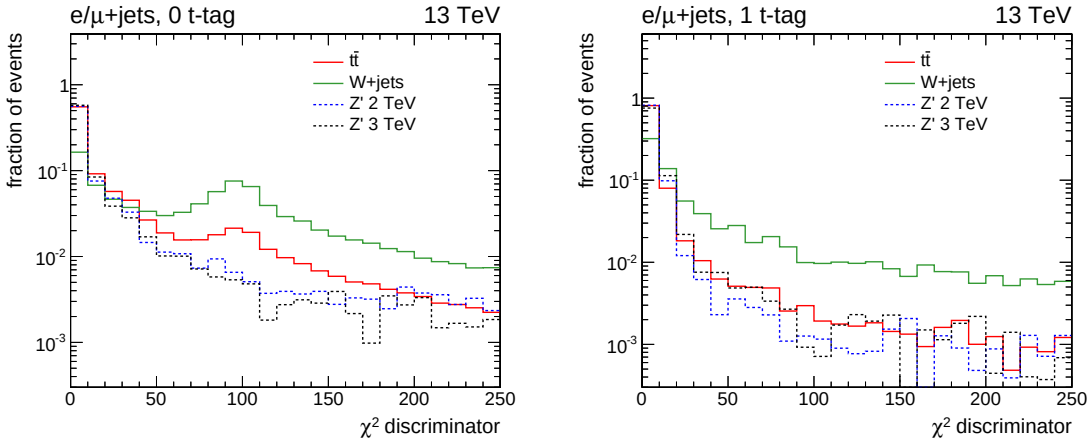


Figure 5.9: Mass of the reconstructed hadronic top quark for the correct $t\bar{t}$ hypothesis (defined by ΔR-matching between generated particles and reconstructed objects) in the μ + jets channel. Events without (with) a t-tagged jet are shown on the left (right). The red line in each plot corresponds to the fit to a gaussian function. Distributions are shown for SM $t\bar{t}$ production (top) and a 1%-width Z′ boson with $M_{Z'} = 2$ TeV (center) and $M_{Z'} = 3$ TeV (bottom).



(a) (b)

Figure 5.10: Distribution of the χ^2 discriminator for events passing the ℓ + jets preselection. Plots are shown separately for the events in the 0 t-tag (a) and 1 t-tag (b) categories. Each distribution is normalized to unity.

The values used in the χ^2 discriminator for events without (with) a t-tagged jet are $\bar{m}_{\text{top}}^{\text{lep}} = 175\,(175)$ GeV, $\sigma_M^{\text{had}} = 19\,(19)$ GeV, $\bar{m}_{\text{top}}^{\text{had}} = 177\,(173)$ GeV and $\sigma_M^{\text{had}} = 16\,(15)$ GeV. Figure 5.10 shows a comparison of the χ^2 discriminator obtained in the 0 t-tag and 1 t-tag categories for different SM backgrounds and signal hypotheses. As expected, this distribution peaks at low values for true $t\bar{t}$ events (SM $t\bar{t}$ and BSM $t\bar{t}$ resonances), whereas non-$t\bar{t}$ backgrounds show a larger tale in the high-χ^2 region. The peak around $\chi^2 \simeq 100$ in the 0 t-tag category for the W + jets and $t\bar{t}$ simulation arises from events with exactly two AK4 jets and only one jet assigned to the hadronic top quark; since the mass of this small-radius jet peaks at zero, like the leptonic term of the χ^2 discriminator, the full χ^2 discriminator assumes values around $\chi^2 \simeq 100$. In the case of SM $t\bar{t}$ events this can occur when one or more quarks from the hadronic top quark decay are not reconstructed as jet candidates, due to the acceptance cuts on the jets.

Finally, Figure 5.11 shows the relative difference between the mass of the reconstructed $t\bar{t}$ pair and the mass of the generated $t\bar{t}$ pair for the correct $t\bar{t}$ hypothesis. For the latter, the expected resolution of the reconstructed $t\bar{t}$ mass ranges from 6% to 8% for both SM $t\bar{t}$ production and different signal mass hypotheses (up to $M_{Z'} = 4$ TeV). A slight improvement in the expected $M_{t\bar{t}}$ resolution is observed as the mass of the signal hypothesis increases. Similar values for the mass resolution are found in 0 t-tag and 1 t-tag events.

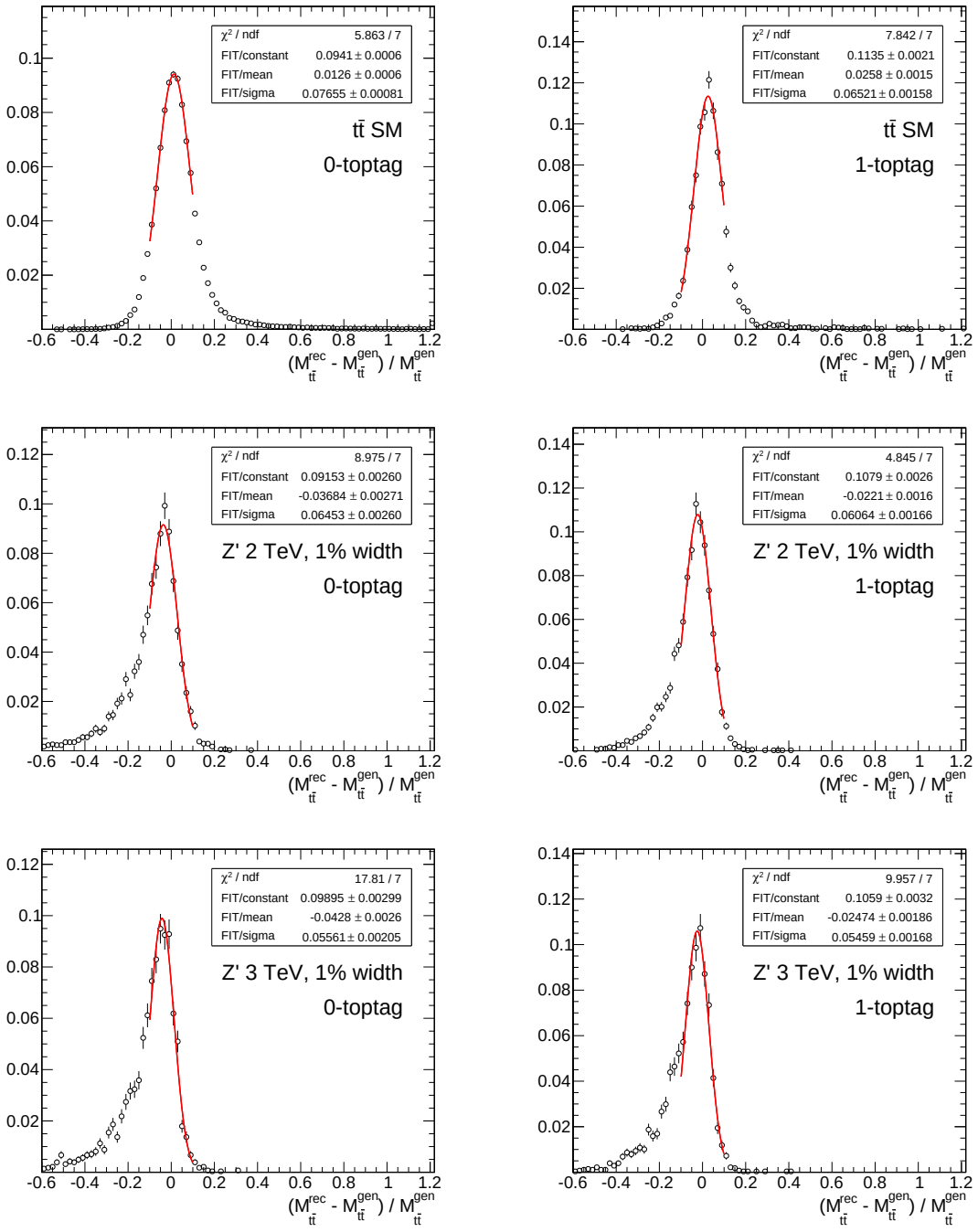


Figure 5.11: Relative difference between reconstructed and generated mass of the $t\bar{t}$ system for the correct $t\bar{t}$ hypothesis (defined by ΔR-matching between generated particles and reconstructed objects) in the μ+jets channel. Events without (with) a t-tagged jet are shown on the left (right). The red line in each plot corresponds to the fit to a gaussian function. Distributions are shown for SM $t\bar{t}$ production (top) and a 1%-width Z′ boson with $M_{Z'} = 2$ TeV (center) and $M_{Z'} = 3$ TeV (bottom).

5.3.4 Final event selection and categorization

An upper cut on the χ^2 discriminator of the best $t\bar{t}$ hypothesis is applied to define the $\ell+$jets signal region. This requirement has the effect of reducing the contribution of non-$t\bar{t}$ backgrounds in the SR and enhance the analysis sensitivity.

The specific threshold used for the χ^2 cut is determined by estimating the expected 95% CL exclusion limits on $\sigma \times$ BR for a narrow-width Z′ signal. This optimization method is also used to determine the best categorization for the final event sample; as already described in Section 4.8, the analysis sensitivity can be maximized by splitting the final event sample in exclusive categories characterized by different background content and signal-over-background ratios. Figure 5.12 shows the expected limits on $\sigma \times$ BR, combined for the $\mu+$jets and $e+$jets channels, for different choices of the χ^2 upper threshold and the event categorization. Based on these tests, the final sensitivity is expected to be mildly dependent on the χ^2 threshold, for values of the latter between 10 and 100; we choose an intermediate threshold at $\chi^2_{\text{cut}} = 30$; this is done to balance the level of event statistics in both the SR ($\chi^2 < 30$) and the control regions defined by $\chi^2 > 30$, which are used in the final background estimation. Figure 5.12(b) shows how the use of different categorizations can impact the analysis sensitivity. The categories tested are defined by splitting events by the number of b-tagged and t-tagged jets. A clear gain in sensitivity is expected using any of these categorizations compared to simply using one inclusive category. Optimal performance is reached when events are separated based on the number of t-tagged jets (zero or at least one); splitting events without a t-tagged jet based on the number of b-tags (zero or at least one) leads to a further improvement of the expected limits, particularly in the low-mass region ($M_{\text{Z}'} < 1.5$ TeV). Further slicing of the SR categories does not improve significantly the expected limits. As a result, the final event categorization corresponds to the one designed for the Run-1 version of this analysis and it is based on three exclusive samples: events with a t-tagged jet (`T1` category), events without a t-tagged jet and with at least one b-tagged jet (`T0B1` category) and events without any t-tagged or b-tagged jets (`T0B0` category). This categorization is applied separately for the $\mu+$jets and $e+$jets channels, for a total of six SR categories.

Signal efficiencies

Figure 5.13 shows the overall signal efficiencies for the three final categories in the $\ell+$jets SR. For each signal MC sample, these efficiencies are calculated over all generated events, which include all possible top quark decay modes. Due to the tight acceptance requirements on the t-tagged jet, most importantly $p_T > 500$ GeV, the 1-t-tag category reaches a plateau efficiency only for signal mass hypotheses with $M_X \gtrsim 2$ TeV. The highest signal efficiencies are obtained for $\text{Z}' \to t\bar{t}$ signals with narrow decay width ($\Gamma_{\text{Z}'}/M_{\text{Z}'} = 1\%$); the efficiencies for signal models with larger widths (10%- and 30%-width Z′ bosons and KK gluon resonances) are slightly lower, due to the higher rate of off-shell production expected for these signals.

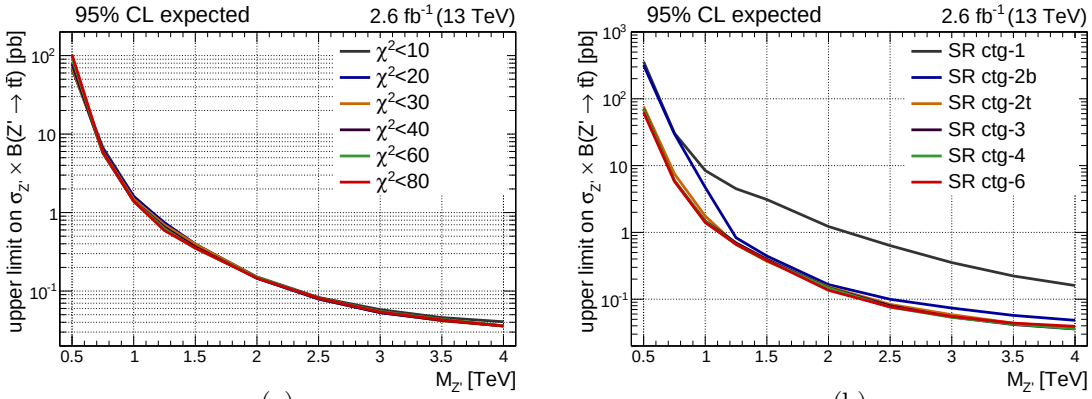


Figure 5.12: Expected 95% CL limits on $\sigma \times \mathrm{BR}$ for a narrow-width $\mathrm{Z}' \to t\bar{t}$ signal, as a function of the signal mass hypothesis. Different choices for the χ^2 cut threshold (a) and the final event categorization (b) are compared. Limits shown in (a) use the "ctg-3" categorization in (b); limits shown in (b) are obtained with $\chi^2 < 30$. Legend for categorizations shown in (b): "ctg-1": inclusive sample; "ctg-2b": no b-tags, at least one b-tag; "ctg-2t": no t-tags, one t-tag; "ctg-3": no t-tags and no b-tags, no t-tags and at least one b-tag, one t-tag; "ctg-4": no t-tags and no b-tags, no t-tags and at least one b-tag, one t-tag and no b-tags, one t-tag and at least one b-tag; "ctg-6": no t-tags and no b-tags, no t-tags and exactly one b-tag, no t-tags and at least two b-tags, at least one t-tag and no b-tags, at least one t-tag and exactly one b-tag, at least one t-tag and at least two b-tags.

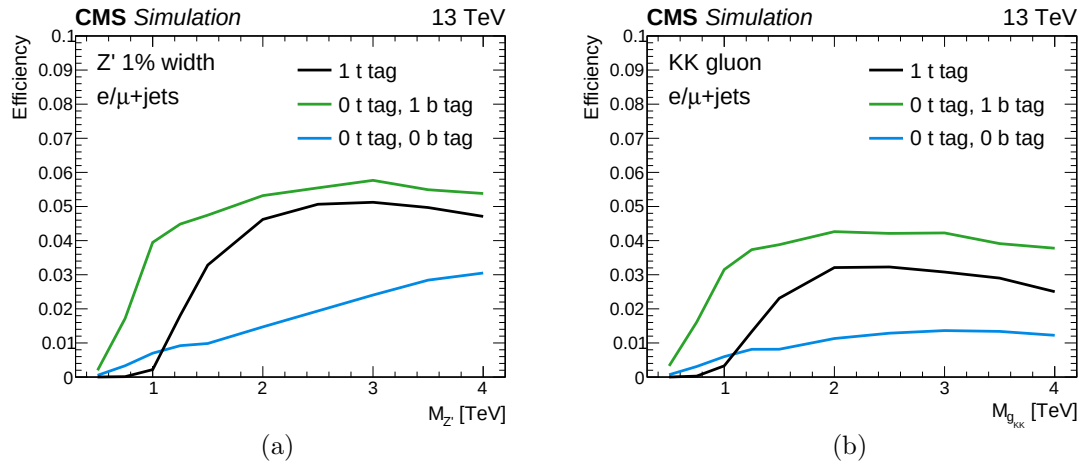


Figure 5.13: Signal efficiencies in the ℓ + jets SR categories as a function of the signal mass hypothesis: (a) Z' boson with 1% width, (b) KK gluon resonance. Each efficiency is calculated on the full generated sample, which includes all possible top quark decay modes. The efficiencies for the μ + jets and e + jets channels are added together.

5.4 Efficiency studies in data and simulated events

After defining the physics objects and event selection of this search, we study the efficiency of the different selection cuts used in the ℓ + jets analysis. We make use of control samples, orthogonal to the ℓ + jets SR, to measure these efficiencies in both data and MC events; when necessary, the measurements in data are used to determine the efficiency corrections to be applied to the simulation and the size of their uncertainty. The selection cuts studied in this section include the lepton identification and isolation cuts, the trigger selection and the jet b-tagging and t-tagging requirements.

5.4.1 Lepton identification and isolation requirements

The selection applied on leptons in the ℓ + jets analysis, aside from acceptance cuts on transverse momentum and pseudorapidity, consists of two independent requirements. The first pertains to the lepton ID criteria, whereas the second selection is concerned with further distinguishing non-prompt leptons from leptons originating directly from the hard-scattering process, for example in a semileptonic top quark decay. This second stage of the lepton selection can include a cut on lepton isolation or the lepton 2D-cut criterion introduced in Section 4.4.

This section describes studies performed with simulated events in order to determine the optimal lepton ID and isolation criteria for this analysis and quantify their performance. The optimization is based on maximizing the efficiency for selecting charged leptons coming from the semileptonic decay of a top quark and, at the same time, minimizing the mis-identification rate for non-prompt leptons.

Only reconstructed leptons matched to a generated lepton from a top quark decay are considered for the measurement of the identification and isolation efficiencies on prompt leptons in simulated events. A reconstructed lepton is said to be matched to a generator-level lepton from a $t \to bW \to b\ell\nu$ decay if all the following conditions are met:

- upper cut on the ΔR-distance between reconstructed and generator-level lepton: $\Delta R(\ell_{\text{RECO}}, \ell_{\text{GEN}}) < 0.1$;
- upper cut on the relative difference in p_T between reconstructed and generator-level lepton: $|p_T^{\ell_{\text{RECO}}} - p_T^{\ell_{\text{GEN}}}|/p_T^{\ell_{\text{GEN}}} < 0.3$;
- same electric charge.

Leptons not satisfying the above matching requirements are used to estimate the fake rate (efficiency for non-prompt leptons) of a given identification or isolation selection.

In the rest of this section, we use the term "signal efficiency" to refer to the efficiency of a given algorithm for prompt (GEN-matched) leptons, whereas the term "background efficiency" is used to indicate the efficiency of a given algorithm for non-prompt leptons; in this context, the term "background rejection" corresponds to one minus the background efficiency. The signal efficiencies for leptons coming from high-p_T top quarks are estimated using simulated events for the SM $t\bar{t}$ production and the production of Z′ boson decaying to a $t\bar{t}$ pair, with masses of 1 TeV, 2 TeV and 3 TeV and a relative width of 1%; for the fake rate estimates, we take as a reference the MC simulation for QCD-multijet production, which is expected to be the main SM process leading to the production of non-prompt leptons in the ℓ + jets final state (before any isolation-based requirement).

In order to properly estimate the lepton efficiencies expected in the ℓ + jets analysis, each efficiency is measured only for events passing the jet and $\not{E}_T$ requirements applied in the

μ + jets (e + jets) preselection. As described in Section 5.3.2, this implies the use of slightly different cuts in the two channels.

5.4.1.1 Muon ID

We consider three different options for the muon ID criteria, known as Loose-ID, Medium-ID and Tight-ID. The muon candidates used to measure these ID efficiencies are required to have $p_T^\mu > 50$ GeV and $|\eta^\mu| < 2.1$.

Figure 5.14 shows the signal efficiencies and background rejection estimates for the three algorithms, as a function of muon p_T, muon η and number of primary vertices. Based on these MC estimates, the Loose-ID gives the best signal efficiency, but also the highest fake rate, among the three methods. The Medium-ID gives a slightly lower signal efficiency compared to the Loose-ID, but a better signal efficiency and the same background rejection of the Tight-ID. This suggests that using the Medium-ID selection allows to obtain the best signal efficiency for the maximum background rejection in the μ + jets channel.

The efficiency of the Medium ID has been measured directly in collisions data using the tag-and-probe method in a sample of $Z \to \mu^+\mu^-$ events [75]. The corresponding data/MC SFs, given in Table 5.2 for separate bins of the muon p_T and η, are used to correct the simulation in the ℓ + jets analysis.

Table 5.2: Data/MC corrections for the efficiency of the muon identification requirements, in separate bins of the muon p_T and η.

$\varepsilon_{\text{Data}}/\varepsilon_{\text{MC}}$	$\lvert\eta^\mu\rvert < 0.9$	$0.9 \leqslant \lvert\eta^\mu\rvert < 1.2$	$1.2 \leqslant \lvert\eta^\mu\rvert < 2.1$
50 GeV $< p_T^\mu \leqslant$ 60 GeV	0.988 ± 0.010	0.992 ± 0.010	0.991 ± 0.010
60 GeV $< p_T^\mu \leqslant 120$ GeV	0.991 ± 0.010	0.992 ± 0.011	0.996 ± 0.012
$p_T^\mu > 120$ GeV	0.991 ± 0.020	0.992 ± 0.022	0.996 ± 0.024

5.4.1.2 Electron ID

For the electron ID selection, the following options are considered [77]:

- cut-based identification method, described in Section 5.3.1, for three working points ("loose", "medium" and "tight");
- non-triggering MVA ID: this ID method is based on a single discriminator obtained from a multivariate analysis of multiple parameters determined in the electron reconstruction; this selection is tested for both a "loose" and a "tight" working point;
- High Energy Electron Pairs (HEEP) ID [144]: the HEEP ID corresponds to a modified version of the cut-based ID, optimized for high-p_T electrons ($p_T > 200$ GeV).

The electron candidates used to measure these ID efficiencies are required to have $p_T^e > 50$ GeV and $|\eta^e_{\text{SC}}| < 2.5$.

Figure 5.15 shows the signal efficiencies and background rejection estimates for the cut-based and non-triggering MVA ID criteria, as a function of electron p_T, electron supercluster η and number of primary vertices. By comparing the MC efficiencies of the different algorithms, the following picture emerges:

- if one favors the optimization of the signal efficiency, the best option is represented by the loose working point of the non-triggering MVA ID, which outperforms the loose cut-based ID in terms of both signal efficiency and background rejection;
- in terms of background rejection, the best performance is obtained using the tight working point of the cut-based ID.

In order to ensure the maximum level of QCD rejection in the $e+$jets, we choose the tight working point of the cut-based ID as the method to identify electron candidates in the analysis.

The efficiency of this electron ID has been measured in a $Z \to e^+e^-$ sample using a tag-and-probe method [77]. The corresponding data/MC SFs, reported in Table 5.3, are used to reweight simulated events in the $\ell+$jets analysis in order to account for differences between the electron ID efficiencies in data and MC.

Table 5.3: Data/MC corrections for the efficiency of the electron identification cuts, binned with respect to the electron p_T and supercluster η.

$\varepsilon_{\text{Data}}/\varepsilon_{\text{MC}}$	50 GeV $< p_T^e \leqslant$ 200 GeV	$p_T^e >$ 200 GeV
$\lvert\eta^e_{\text{SC}}\rvert < 0.8$	0.986 ± 0.006	0.986 ± 0.012
$0.8 \leqslant \lvert\eta^e_{\text{SC}}\rvert < 1.4442$	0.981 ± 0.006	0.981 ± 0.012
$1.4442 \leqslant \lvert\eta^e_{\text{SC}}\rvert < 1.566$	0.961 ± 0.020	0.961 ± 0.040
$1.566 \leqslant \lvert\eta^e_{\text{SC}}\rvert < 2.0$	0.996 ± 0.009	0.996 ± 0.018
$2.0 \leqslant \lvert\eta^e_{\text{SC}}\rvert < 2.5$	1.011 ± 0.012	1.011 ± 0.024

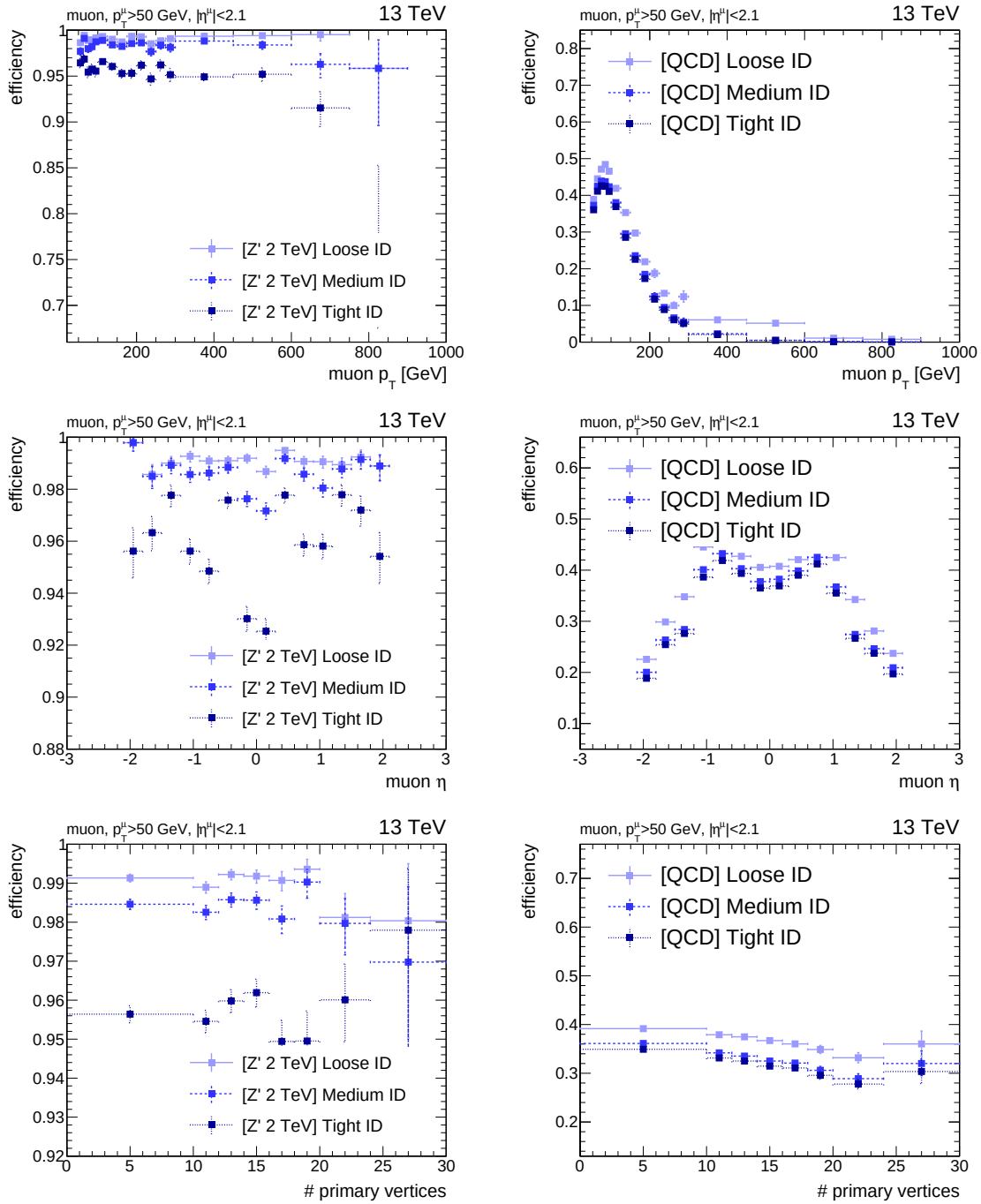


Figure 5.14: Efficiency for different muon identification criteria: (left) signal efficiency in MC for the production of a $Z' \to t\bar{t}$ signal with $M_{Z'} = 2$ TeV and $\Gamma_{Z'}/M_{Z'} = 1\%$, (right) background efficiency in MC for QCD-multijet production. Efficiencies are shown as a function of (a) muon p_T, (b) muon η and (c) number of primary vertices in the event.

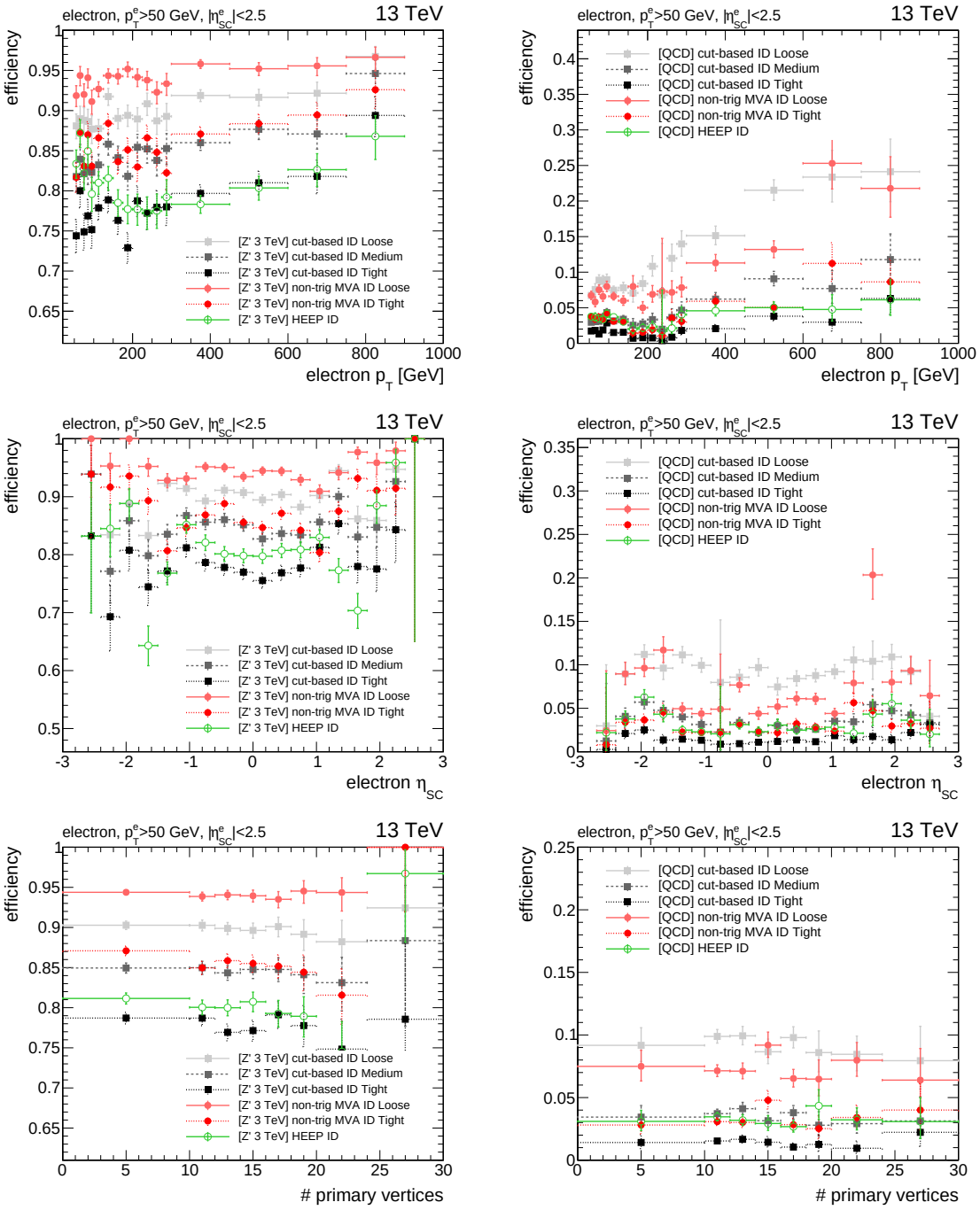


Figure 5.15: Efficiency for different electron identification criteria: (left) signal efficiency in MC for the production of a $Z' \to t\bar{t}$ signal with $M_{Z'} = 3$ TeV and $\Gamma_{Z'}/M_{Z'} = 1\%$, (right) background efficiency in MC for QCD-multijet production. Efficiencies are shown as a function of (a) electron p_T, (b) electron supercluster η and (c) number of primary vertices in the event.

5.4.1.3 Lepton isolation

The second stage of the lepton selection is related to a requirement on lepton isolation, a property which is expected to characterize prompt leptons, as opposed to non-prompt ones. In the case of the decay of high-p_T top quarks, the situation is complicated by the fact that the prompt lepton from the W boson decay is expected to be poorly isolated at the reconstruction level, due to its vicinity in space to the hadronic jet originated by the b-quark. For this reason, standard fixed-cone isolation algorithms have been found to be not suited for searches involving high-p_T ($p_T^t \gtrsim 400$ GeV) top quarks.

In the following we compare three algorithms:

- PF-based fixed-cone isolation: this is the standard method to identify prompt leptons in most CMS physics analyses, based on the quantity defined in Equation (3.1); for muons (electrons), isolation deposits are calculated using a cone of size $R = 0.4$ ($R = 0.3$) around the direction of the lepton momentum; the pileup contribution is subtracted using the $\Delta\beta$ correction [145];

- PF-based mini-isolation [146]: this method is similar to the standard PF-isolation, with the only difference that the isolation cone-size is parameterized as a function of the lepton p_T, as a way to improve the efficiency for leptons originating from high-p_T top quarks. The cone-size parameterization considered here is the following:

$$R_{\text{iso}}(p_T^\ell) = \frac{K_T}{p_T^\ell} \quad \text{and} \quad R_{\text{min}} < R_{\text{iso}}(p_T^\ell) < R_{\text{max}} \tag{5.3}$$

  where $K_T = 10$ GeV, $R_{\text{min}} = 0.05$ and $R_{\text{max}} = 0.2$; the pileup contribution is subtracted using the $\Delta\beta$ correction;

- lepton 2D-cut [118]: this selection cut, already used in the Run-1 version of this analysis, is given by the logical OR of two requirements

$$\Delta R_{\text{min}}(\ell, j) > \delta R^{\ell j} \quad || \quad p_{T,\text{rel}}(\ell, j) > p_{T,\text{rel}}^{\ell j} \tag{5.4}$$

  where $\Delta R_{\text{min}}(\ell, j)$ is the minimum ΔR-distance between the lepton candidate and all AK4 jets with $p_T > p_{T,\text{min}}^j$ and $|\eta^j| < \eta_{\text{max}}^j$ and $p_{T,\text{rel}}(\ell, j)$ is the transverse momentum of the lepton with respect to the axis of ΔR-nearest AK4 jet with $p_T > p_{T,\text{min}}^j$ and $|\eta^j| < \eta_{\text{max}}^j$. Based on this definition, the lepton 2D-cut depends on four independent parameters: $\delta R^{\ell j}$, $p_{T,\text{rel}}^{\ell j}$, $p_{T,\text{min}}^j$ and η_{max}^j.

Figure 5.16 shows a comparison of the three algorithms using *receiver operating characteristic* (ROC) curves for signal efficiency versus background rejection. These efficiencies are measured using muon candidates passing $p_T^\mu > 50$ GeV, $|\eta^\mu| < 2.1$ and the Medium-ID requirement, and electron candidates passing $p_T^e > 50$ GeV, $|\eta_{\text{SC}}^e| < 2.5$ and the tight working point of the cut-based ID. The ROC curves suggest that the standard fixed-cone lepton isolation is clearly ineffective for searches involving high-p_T top quarks: acceptable signal efficiencies can only be reached with extremely loose cuts that lead to a very poor rejection power for non-prompt leptons. On the other hand, the lepton 2D-cut and PF-based mini-isolation can reach the same level of signal efficiency, but the former method is much more effective than the latter in rejecting fake leptons from QCD production for the event selection considered in this study. These features are shown in Figure 5.16 for a $Z' \to t\bar{t}$ MC with $M_{Z'} = 3$ TeV, but the same conclusions hold when considering the $t\bar{t}$ MC and different signal hypotheses. As a consequence, the lepton 2D-cut remains the preferred option to complement the lepton ID requirements in the analysis.

Lepton 2D-cut tuning

We also consider possible improvements in the parameterization of the lepton 2D-cut. Figure 5.17 shows the ROC curves for different parameterizations of the lepton 2D-cut; specifically, every curve is defined by a certain choice of the $\delta R^{\ell j}$, $p^{j}_{T,\mathrm{min}}$ and η^{j}_{max} parameters, and every point corresponds to a different value for $p^{\ell j}_{T,\mathrm{rel}}$. These MC efficiencies suggest that the background rejection of the lepton 2D-cut is mostly sensitive to the p_T threshold on the jets used in the calculation of $\Delta R_{\mathrm{min}}(\ell, j)$ and $p_{T,\mathrm{rel}}$; a looser jet selection (lower jet p_T threshold) leads to an improvement in the background rejection, without compromising the efficiency for prompt leptons, which can be adjusted by varying the $p^{\ell j}_{T,\mathrm{rel}}$ threshold. Compared to the jet p_T threshold, the η restriction on the jets has a negligible impact on the lepton 2D-cut response, as long as it is looser than the one used for the lepton candidate; nevertheless, to maximize the background rejection of the cut, we apply a loose requirement on the jets η, given by $|\eta^{j}| < 3.0$.

Based on these MC studies, the optimal working point for the lepton 2D-cut is given by the following choice of its parameters: $\delta R^{\ell j} = 0.4$, $p^{\ell j}_{T,\mathrm{rel}} = 20$ GeV, $p^{j}_{T,\mathrm{min}} = 15$ GeV and $\eta^{j}_{\mathrm{max}} = 3.0$. Finally, Figure 5.18 shows the efficiency of the working point chosen for the lepton 2D-cut, as a function of the lepton p_T, the lepton η and the number of primary vertices. These results show that the lepton 2D-cut efficiency has a very limited dependence on pileup and on the position of the reconstructed lepton.

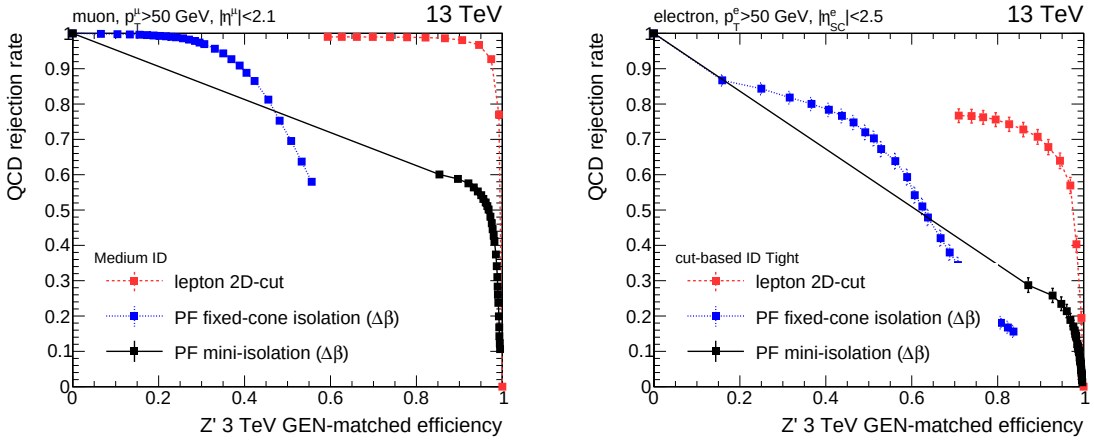


Figure 5.16: ROC curves for signal efficiency versus background (QCD) rejection in simulated events for PF-based fixed-cone isolation (blue), PF-based mini-isolation (black) and lepton 2D-cut (red): (left) efficiency for muon candidates, (right) efficiency for electron candidates. The lepton signal efficiency is estimated for a $Z' \to t\bar{t}$ signal with $M_{Z'} = 3$ TeV and 1% width.

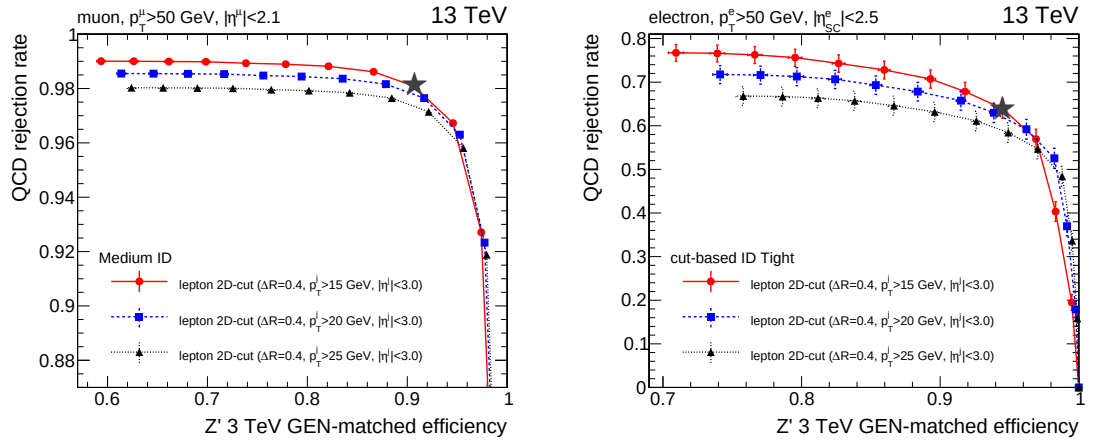


Figure 5.17: ROC curves for signal efficiency versus background (QCD) rejection in simulated events, for different choices of the lepton 2D-cut parameters: (left) efficiency for muon candidates, (right) efficiency for electron candidates. The lepton signal efficiency is estimated for a $Z' \to t\bar{t}$ signal with $M_{Z'} = 3$ TeV and 1% width. The working point used in the ℓ+jets analysis is marked with a star symbol.

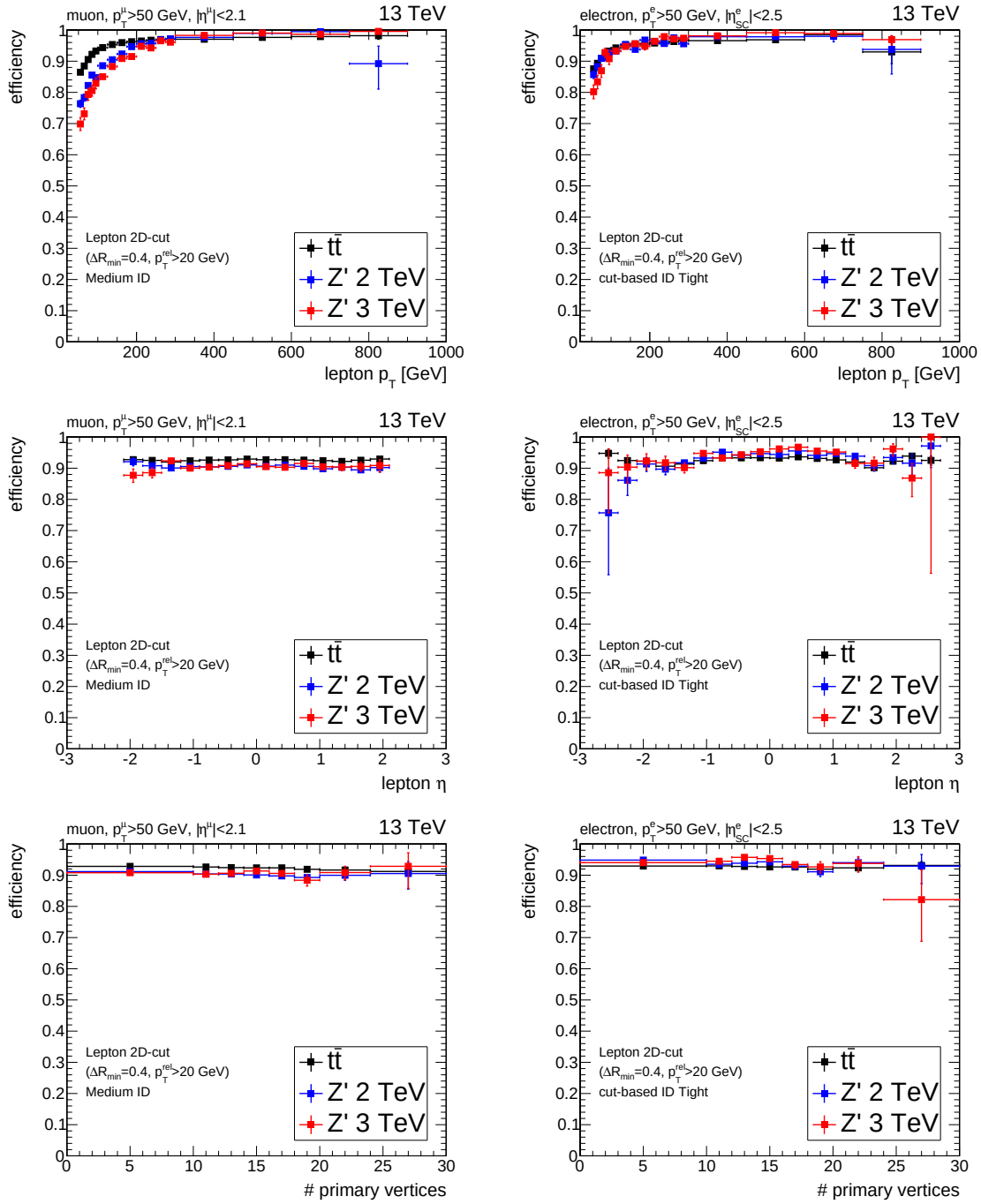


Figure 5.18: Efficiency for the lepton 2D-cut working point used in the ℓ + jets analysis: (left) efficiency for muon candidates, (right) efficiency for electron candidates. Efficiencies are shown for signal leptons (lepton matched to a GEN top quark) as a function of (top) lepton p_T, (center) lepton η and (bottom) number of primary vertices.

5.4.1.4 Lepton 2D-cut efficiency in $Z \to \ell\ell$ control sample

As described in the previous sections, the lepton 2D-cut is the selection used in the analysis to distinguish prompt and non-prompt leptons. In order to measure the efficiency of this cut directly in data, we select a control region dominated by $Z(\to \ell^+\ell^-)$ + jets production. The tag-and-probe method [75, 77] is used in this sample to measure the efficiency of the lepton 2D-cut.

The dimuon and dielectron control regions are defined as follows. Each event is required to have exactly two leptons, with the same lepton flavor and opposite electric charge, passing the p_T and η cuts used for the lepton candidates of the ℓ + jets analysis (see Section 5.3.1). Among the two leptons in the events, one has to be identified as the "tag" lepton; the latter has to satisfy tight identification requirements and a tight cut on the standard fixed-cone isolation. The other lepton is identified as "probe" lepton and it is only required to satisfy the same identification cuts applied on the lepton candidates of the ℓ + jets analysis. The triggers used in both the ee and $\mu\mu$ control regions require the presence of one isolated lepton at trigger level. The tag lepton is required to be matched to the online object triggering the event; this trigger matching is satisfied if the two objects have an angular separation lower than $\Delta R = 0.1$. This matching requirement on the tag is intended to minimize the probability of identifying a non-prompt lepton as tag and to avoid any bias in the efficiency measurement performed on the probe lepton. In order to mimic the jet requirements of the ℓ + jets analysis, we apply the same kinematic cuts on jets in the dilepton control sample: this implies $p_T^{\text{jet-1}} > 150\,(250)$ GeV and $p_T^{\text{jet-2}} > 50\,(70)$ GeV in the $\mu\mu$ (ee) CR. No cut on the missing transverse energy is applied in the two CRs, as no real missing transverse energy is expected in a sample dominated by DY production; on the other hand, in order to further mimic the phase space selected in ℓ + jets analysis, the tag lepton in the dilepton CRs is required to pass the same requirements applied on $\not{E}_T$ in the ℓ + jets SRs; as a consequence, we require $p_T^{\text{TAG}} > 50$ GeV and $p_T^{\text{TAG}} + p_T^{\text{PRO}} > 150$ GeV in the $\mu\mu$ CR and $p_T^{\text{TAG}} > 120$ GeV in the ee CR. Figure 5.19 shows data/MC distributions for the $Z(\to \ell\ell)$ + jets samples defined above. Reasonable agreement between data and simulation is observed in all the distributions considered in these two CRs, which are both fully dominated by DY+jets production.

These control regions are used to measure the efficiency of the lepton 2D-cut on the probe lepton. For both the dimuon and dielectron mass spectra, the interval from 60 GeV to 130 GeV is used in the fits. The probability function used to fit the signal component is given by the convolution of a Breit-Wigner distribution with a crystal-ball function; the use of the latter allows to properly model the low-mass tale generated by leptons losing part of their energy due to photon radiation. The background component of the fit is modeled with an exponential. The dilepton mass spectra for failing and passing probes are fitted independently using a linear combination of the signal and background functions. The efficiency for prompt leptons corresponds to $\varepsilon = N_{\text{pass}}/(N_{\text{pass}} + N_{\text{fail}})$, where N is the signal yield given by the fit.

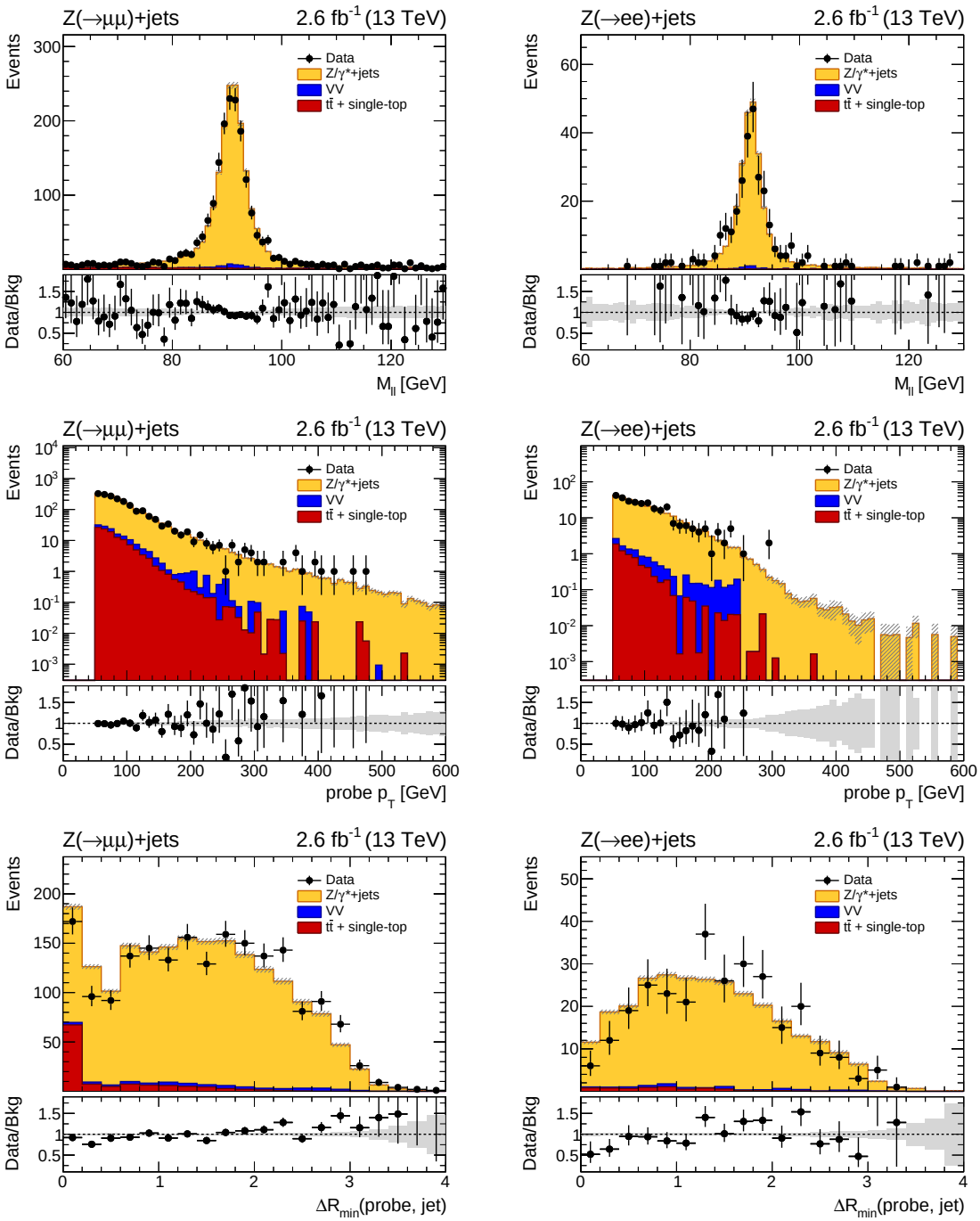


Figure 5.19: Data/MC distributions for the Z($\to \mu\mu$) + jets (left) and Z($\to ee$) + jets (right) control samples: (top) dilepton mass, (center) probe lepton p_T, (bottom) minimum ΔR distance between probe lepton and AK4 jets with $p_T^j > 15$ GeV and $|\eta^j| < 3.0$. The total expected background is normalized to the data and the background error band includes only the MC statistical uncertainty.

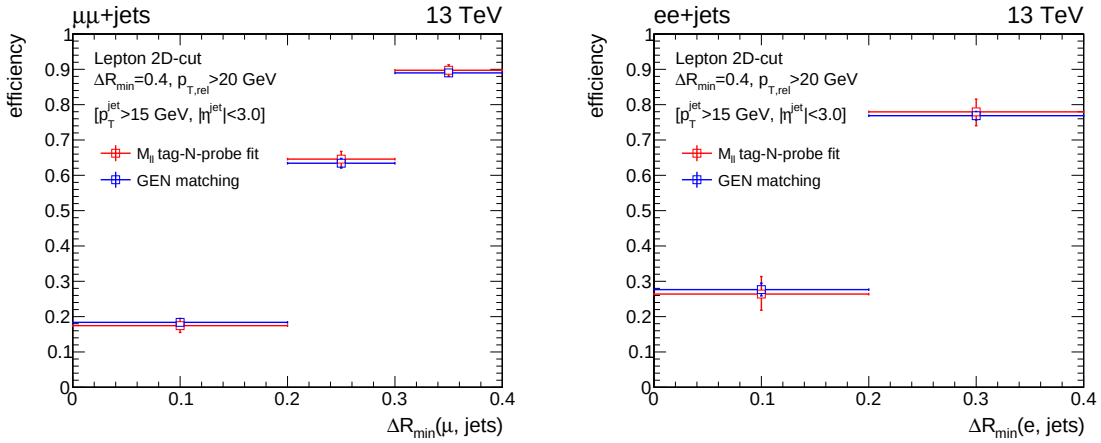


Figure 5.20: Lepton 2D-cut efficiency in simulated events, as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$, comparing the outcome of the tag-and-probe method (red) and the efficiency obtained for a probe lepton matched to a generated lepton (blue): (left) muons, (right) electrons.

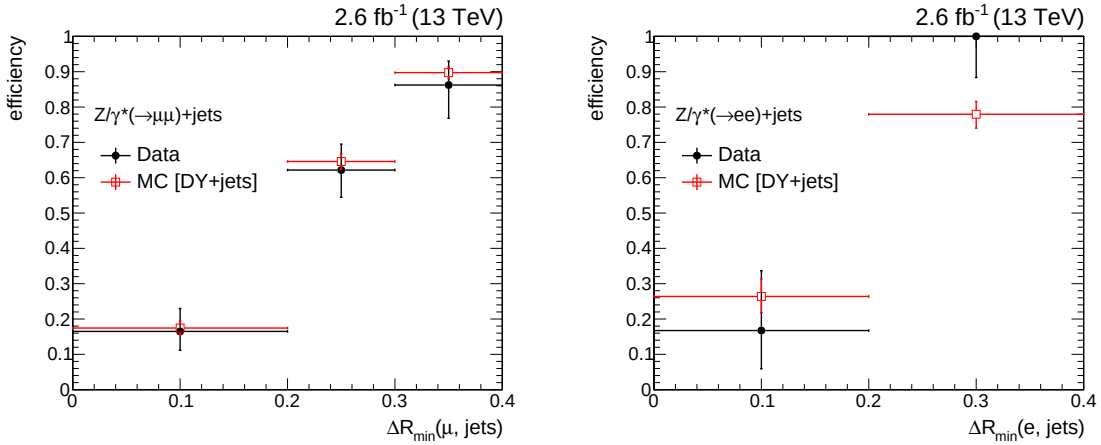


Figure 5.21: Lepton 2D-cut efficiency for data and MC, in the $\mu\mu$ (left) and ee (right) control regions, as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$.

Since the lepton 2D-cut corresponds to a cut on $p_{T,\text{rel}}$ for events with $\Delta R_{\text{min}}(\ell, \text{jets}) < 0.4$, we restrict ourselves in the latter low-ΔR region and measure the lepton 2D-cut efficiency as a function of $\Delta R_{\text{min}}(\ell, \text{jets})$. The efficiency obtained with the tag-and-probe mass fits is validated in simulated events by comparing it to the plain efficiency measured on a probe lepton matched to a generator-level lepton coming from the Z decay. This comparison is displayed in Figure 5.20 and it shows very good agreement between the two measurements, giving confidence in the efficiencies determined with the the tag-and-probe method. Finally, the tag-and-probe efficiencies for data and simulation in the two CRs are shown in Figure 5.21. For both channels, the lepton 2D-cut efficiencies in data and MC are found to be compatible within their statistical uncertainties; therefore, no specific data/MC correction factors for this cut are applied in the ℓ + jets analysis.

5.4.2 High Level Trigger

In this section we study the efficiency of the trigger selections used in the $\mu+$jets and $e+$jets channels. We also measure the performance of these triggers in data and derive data/MC scale factors, which are used to correct the MC predictions in the $\ell+$jets analysis.

5.4.2.1 Trigger efficiency in MC for $\ell+$jets events

Events in $\mu+$jets channel have to pass a trigger requiring an online muon with $p_T > 45$ GeV and $|\eta| < 2.1$. Based on the turn-on curve for the efficiency of the single-muon trigger in MC, shown in Figure 5.22, an optimal trigger efficiency is expected using an offline cut of 50 GeV on the reconstructed muon p_T. This justifies the requirements of with $p_T > 50$ GeV and $|\eta| < 2.1$ on the muon candidate used in the offline analysis (see Section 5.3.1).

The trigger used in the $e+$jets channel is fired if at least an electron with $p_T > 45$ GeV, one jet with $p_T > 200$ GeV and another jet with $p_T > 50$ GeV are reconstructed online. By measuring the turn-on curves of this trigger in simulated events, shown in Figure 5.23 for the electron p_T and the leading-jet p_T, we determine that offline thresholds of 50 GeV for the electron p_T, 250 GeV for the leading jet p_T and 70 GeV for the subleading jet p_T are necessary to reach the trigger plateau efficiency; these cuts are thus the ones applied offline in the $e+$jets analysis.

Figure 5.24 shows the trigger efficiency in simulation for different Z′ mass hypotheses, for the events passing the $\mu+$jets and $e+$jets preselection cuts. It is worth reminding that none of these two triggers makes use of any isolation requirement on the lepton. The efficiency of both HLT selections remains above 90% in the whole mass range considered for this search. The decrease in efficiency of the single-muon trigger from 95% to 90% for high-mass resonances is due to a small dependence of the trigger efficiency on the muon p_T; this dependence is related to the impact of radiative processes, e.g. bremsstrahlung and e^+e^- pair production, on very high-p_T muons, as they interact with the detector material [144]. Compared to the "electron+2-jets" trigger used in the Run-1 version of this analysis, the Run-2 "electron+2-jets" trigger has been improved to retain an optimal efficiency for high-mass $t\bar{t}$ resonances ($M_X > 2$ TeV). The main change in this HLT algorithm with respect to its Run-1

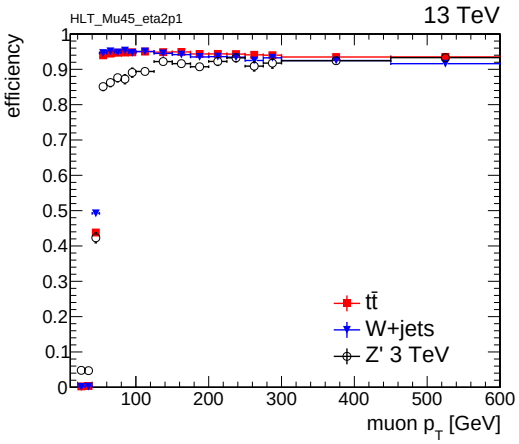
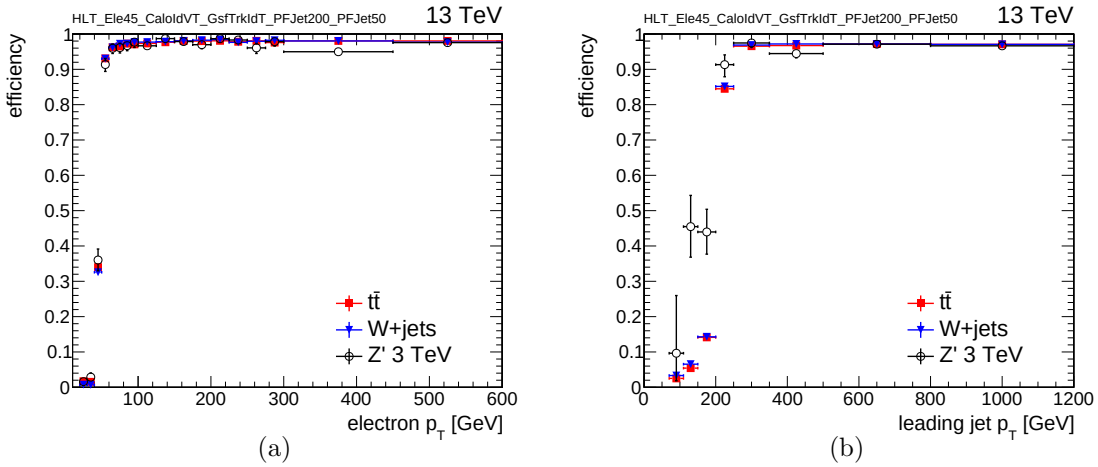


Figure 5.22: Efficiency of the `HLT_Mu45_eta2p1_v*` HLT path, as a function of the offline muon p_T, for events passing the $\ell+$jets preselection. Efficiencies are compared for the $t\bar{t}$ MC, the W + jets MC and a Z′ signal sample.



(a) (b)

Figure 5.23: Efficiency of the $e + 2$ jets trigger used in the electron channel as a function of the offline electron p_T and leading-jet p_T, for events passing the ℓ + jets preselection (without cutting on the variable for which the efficiency is measured). Efficiencies are shown for the $t\bar{t}$ MC, the W + jets MC and a Z′ signal sample.

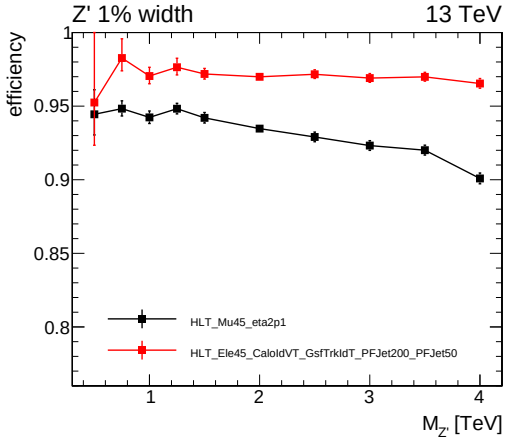


Figure 5.24: Trigger efficiency in the μ + jets and e + jets channels for events passing the corresponding preselection cuts (see Section 5.3.2). The overall HLT efficiencies for different $Z' \to t\bar{t}$ samples are plotted as a function of the signal mass hypothesis.

version is that jet candidates are now cleaned with respect to leptons at trigger level; this allows for an improved reconstruction of final states with boosted semileptonic top quark decays and it results in a higher efficiency for high-mass signals. As a consequence, this electron-based trigger is used standalone, without combining it with other triggers as it was done in Run-1.

5.4.2.2 Trigger efficiency measurements in data

The efficiency of the single-muon trigger has been measured directly in data using the tag-and-probe method in a control sample dominated by $Z \to \mu^+\mu^-$ production [75]. Table 5.4 includes the ratio $\varepsilon_{\text{Data}}/\varepsilon_{\text{MC}}$ of the muon trigger efficiencies measured in data and MC in

Table 5.4: Data/MC corrections for the efficiency of the HLT_Mu45_eta2p1 single-muon trigger, for separate bins of the muon p_T and η.

$\varepsilon_{\text{Data}}/\varepsilon_{\text{MC}}$	$\lvert\eta^\mu\rvert < 0.9$	$0.9 \leqslant \lvert\eta^\mu\rvert < 1.2$	$1.2 \leqslant \lvert\eta^\mu\rvert < 2.1$
50 GeV $< p_T^\mu \leqslant$ 60 GeV	0.974 ± 0.005	0.969 ± 0.006	0.958 ± 0.005
60 GeV $< p_T^\mu \leqslant 120$ GeV	0.969 ± 0.005	0.960 ± 0.006	0.965 ± 0.005
$p_T^\mu > 120$ GeV	0.969 ± 0.010	0.960 ± 0.012	0.965 ± 0.010

different bins of the muon p_T and η; these data/MC SFs are used to correct simulated events in the ℓ + jets analysis.

The performance of the trigger used in the e + jets channel is studied in data using $e\mu$ events. This dilepton control region is defined by the following event selection. Events are triggered by the single-muon HLT path and they are required to have exactly one muon and one electron candidate, as defined in Section 5.3.1; like in the ℓ + jets selection, each lepton candidate has to pass the lepton 2D-cut. The kinematic cuts on jets are the same of the e + jets preselection, i.e. $p_T^{\text{jet-1}} > 250$ GeV and $p_T^{\text{jet-2}} > 70$ GeV. No cut on $\not\!\!E_T$ is applied in this control region, since the previous cuts alone already provide an $e\mu$ sample dominated by $t\bar{t}$ production and the trigger efficiency is not expected to depend on the missing transverse energy. The requirements on electron and jets are the same of the e + jets channel, so the "electron+2-jets" trigger is expected to be fully efficient in this $e\mu$ control sample. A comparison between data and simulation for events in the $e\mu$ control region is given in Figure 5.25. Good agreement is found between data and MC. The control sample is dominated by SM $t\bar{t}$ production ($\simeq 90\%$), with a small contribution ($\simeq 10\%$) from single-top and W + jets events.

These events are used to study the efficiency of the "electron+2-jets" trigger and determine the corresponding data/MC corrections. The efficiencies in data and simulation, and the corresponding data/MC ratios, are shown in Figure 5.25, as a function of some relevant kinematic quantities; the data/MC SFs are determined using the $t\bar{t}$ MC. No dependence on any of the kinematic quantities considered is observed for neither the trigger efficiencies nor the data/MC ratios, within the statistical uncertainty of these measurements. The efficiency ratios are thus fitted with a constant function, as shown in Figure 5.25, to determine the final correction factor. The data/MC scale factor for the efficiency of the electron trigger is estimated to be $\text{SF}_e^{\text{HLT}} = 0.97 \pm 0.02$; this value is used to correct the trigger efficiency of simulated events in the e + jets analysis.

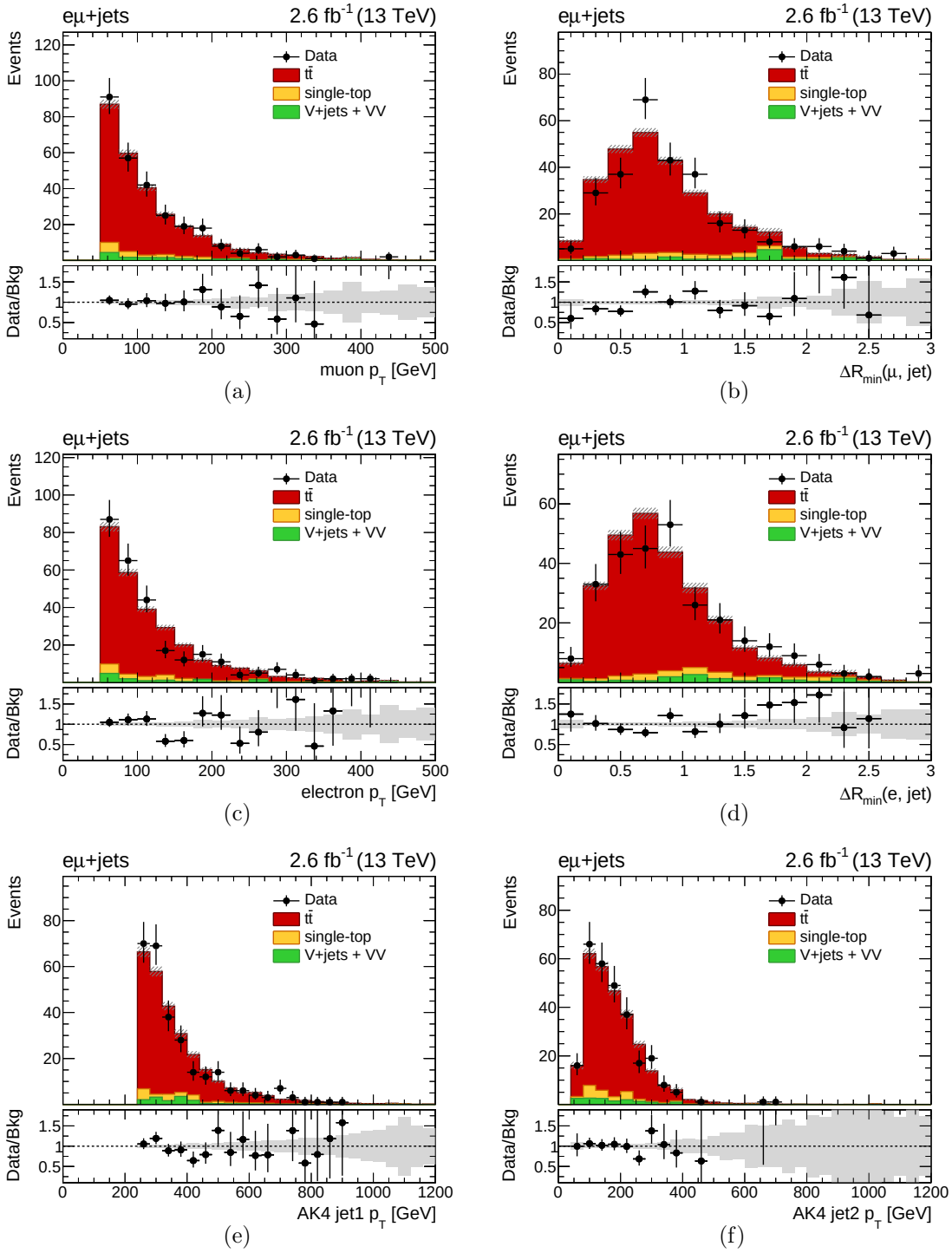


Figure 5.25: Data/MC distributions for the $e\mu$ control sample used to measure the efficiency of the electron trigger: (a) muon p_T, (b) minimum ΔR distance between muon and AK4 jets, (c) electron p_T, (d) minimum ΔR distance between electron and AK4 jets, (e) leading jet p_T, (f) subleading jet p_T. The total expected background is normalized to the data and the background error band includes only the MC statistical uncertainty.

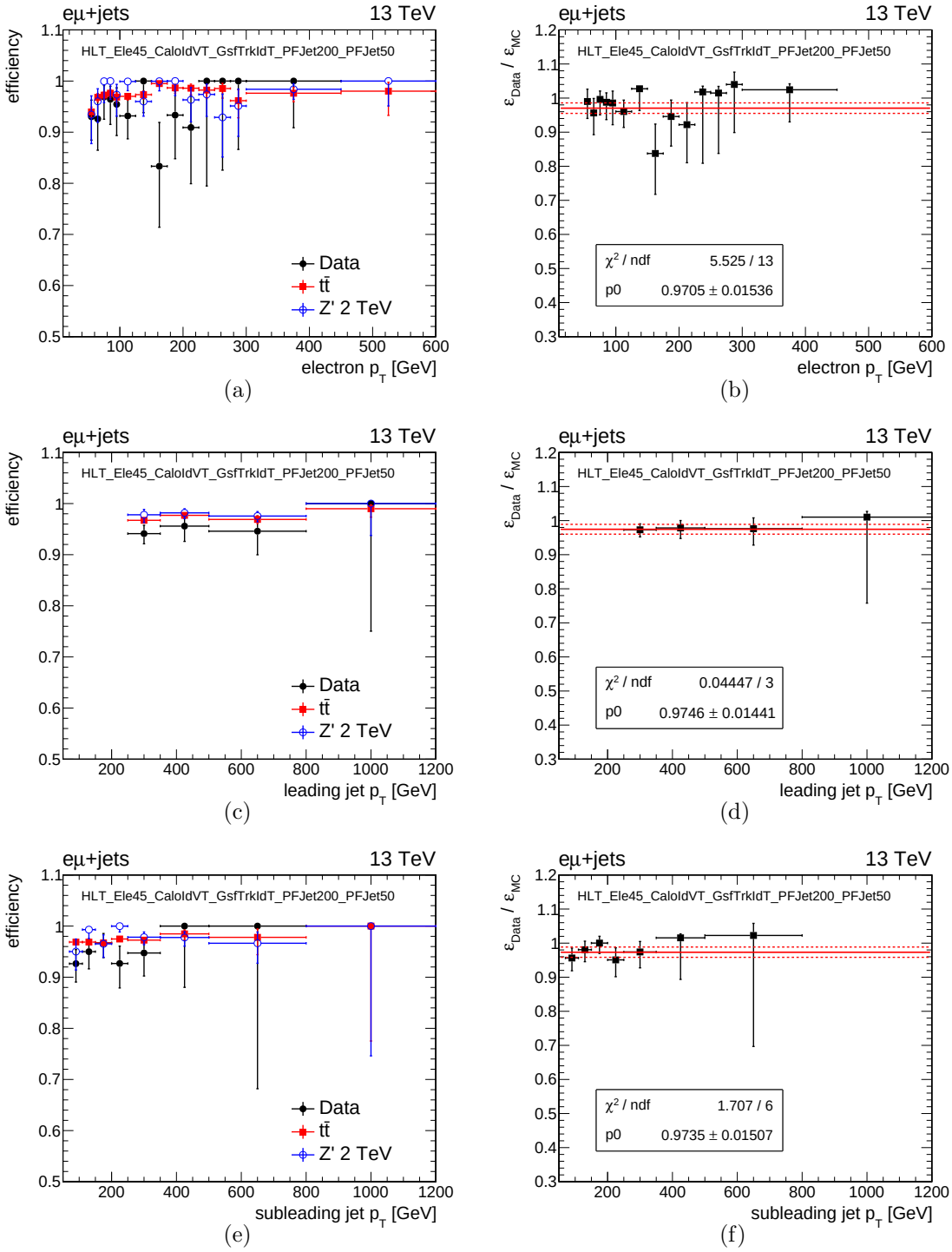


Figure 5.26: Efficiency of the "electron+2-jets" trigger for data and simulated events in the $e\mu$ control sample. For a given kinematic quantity, the trigger efficiency (left) and the corresponding data/MC SF (right) are shown: (a, b) electron p_T, (c, d) leading jet p_T, (e, f) subleading jet p_T.

5.4.3 Jet b-tagging and t-tagging

The use of algorithms to identify jets originating from bottom quark or top quark decays is very important in this analysis to construct event samples dominated by top quark production, with limited contamination due to all the other reducible backgrounds, e.g. W + jets.

Potential discrepancies between data and simulation for the b-tagging and t-tagging efficiencies are accounted for as follows. For both algorithms, MC events are reweighted by the correction factor

$$w = \prod_{\substack{i=1 \\ \text{(tagged)}}} \mathrm{SF}_f(p_T^i\,,\,\eta^i) \prod_{\substack{j=1 \\ \text{(not tagged)}}} \frac{1 - \varepsilon_f^{\mathrm{MC}}(p_T^j) \cdot \mathrm{SF}_f(p_T^j\,,\,\eta^i)}{1 - \varepsilon_f^{\mathrm{MC}}(p_T^j)}\,, \tag{5.5}$$

where the first (second) product loops over the jet candidates passing (not passing) the tagging selection; the subscript f refers to the quark-flavor assigned to a given jet candidate in simulation. The reweighting factor in Equation (5.5) requires, for each jet flavor, the measurement of the tagging efficiency in MC ($\varepsilon_f^{\mathrm{MC}}$) and the data/MC efficiency SF (SF_f); each of these quantities can be determined, in principle, as a function of the jet properties, e.g. jet p_T and η. The reweighting factors for b-tagging and t-tagging are calculated independently.

For the b-tagging correction, the weight is calculated using AK4 jet candidates. Each of the latter is assigned one of three possible flavors: $f = b$, $f = c$ or $f = l$ (light); the jet-flavor assignment for b-tagging in MC events is based on the flavor information of the generated particles clustered in the jet [86].

In the case of t-tagging, only AK8 jet candidates are used to compute the reweghting factor in Equation (5.5). The jet flavor is determined as follows: if there is a top quark at generator-level which lies within a cone of radius $R = 0.8$ around the direction of a given AK8 jet, the latter jet is assigned top flavor ($f = t$); otherwise, the jet is associated to the decay of a gluon or light quark ($f = l$).

In the following we discuss how the different factors needed to calculate the b-tagging and t-tagging corrections are determined.

Jet b-tagging

The b-tagging efficiency in MC events, as a function of the jet p_T, for b-, c- and light-flavor jets are shown in Figure 5.27. The efficiency for true b-jets shows a dependence with respect to jet p_T; it ranges from approximately 65% for jets with $p_T \lesssim 200$ GeV down to to 45% for jets with higher transverse momentum. The efficiency for c-flavor jets in MC lies in a range between 10% and 15% Finally, the b-mistag rate for light-flavor jets varies between less than 1% up to 2%, depending on the jet p_T.

The performance of this b-tagging algorithm has been studied in collisions data at $\sqrt{s} = 13$ TeV by the CMS collaboration [86]. The data/MC SFs for the b-tagging efficiency on true b-jets have been determined, as a function of the jet p_T, in a dijet sample enriched in heavy-flavor content; the relative uncertainty on this correction factor varies from 2% up to 5% in the higher end of the jet p_T spectrum. The b-tagging mis-identification rate for light-flavor jets has been measured in an inclusive multijet sample; the uncertainty on the corresponding data/MC SF amounts to approximately 20%. These corrections are employed in the ℓ + jets analysis to correct the b-tagging efficiencies in simulated events according to Equation (5.5).

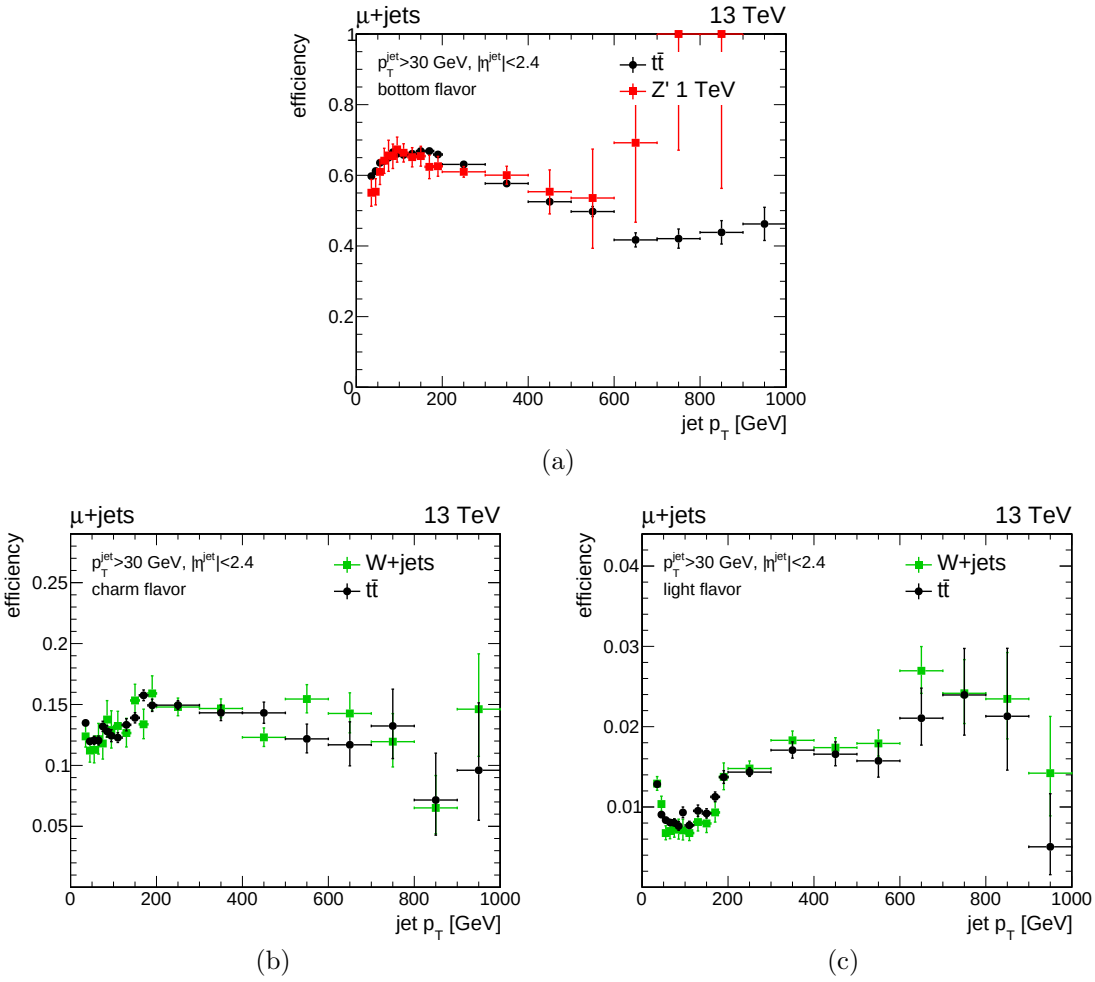


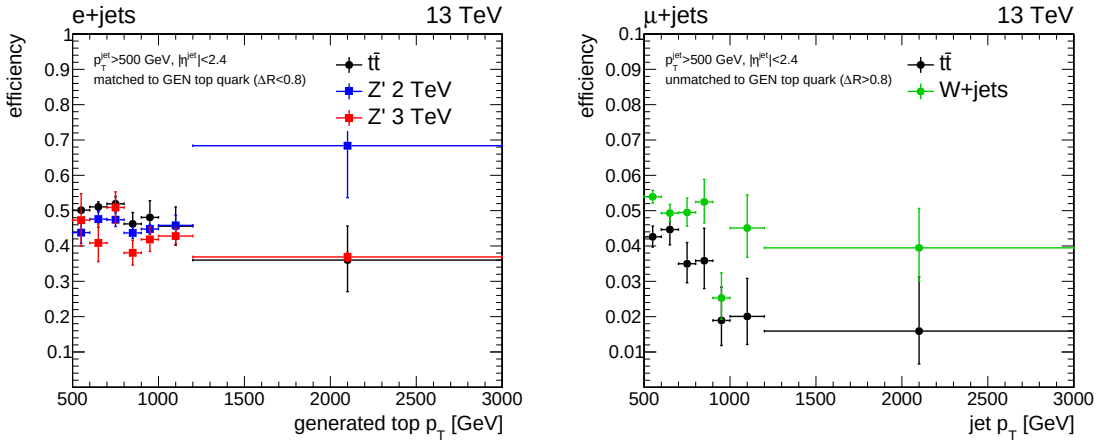
Figure 5.27: MC efficiencies for the b-tagging working point used in the μ + jets analysis. The efficiencies are measured as a function of the AK4 jet p_T in events passing the full event selection described in Section 5.3.2: (a) jets with b flavor, (b) jets with c flavor, (c) jets with light (u, d, s, g) flavor.

Jet t-tagging

The t-tagging efficiencies and t-mistag rate measured in MC events are shown in Figure 5.28 for different simulated processes. Both these efficiencies are measured in MC as a function of the AK8 jet p_T. The efficiency of the t-tagging selection for true t-jets depends mildly on the jet p_T and it ranges between 40% and 50%. The t-mistag rate for light jets is expected to vary between 3% and 6%; the difference, highlighted in Figure 5.28, between the efficiencies in W + jets and $t\bar{t}$ events is due to the different non-top jet flavors dominating in the two cases; additional jets in W + jets production are mostly coming from the decay of quarks, whereas mis-identified jets in $t\bar{t}$ events more often originate from gluons [88].

The following strategy is employed in the analysis regarding the data/MC corrections for the t-tagging efficiencies:

- a data/MC SF for the efficiency of the t-tagging algorithm is not measured directly, but it is introduced as a free nuisance parameter in the analysis; as a consequence, its value is fitted during the limit setting procedure, as detailed in Section 5.5 and



(a) (b)

Figure 5.28: MC efficiencies for the t-tagging selection used in the ℓ + jets analysis, based on the softdrop mass and τ_{32} ratio. The efficiencies are measured in events passing the analysis selection described in Section 5.3.2: (a) t-tagging efficiency in the e + jets channel for AK8 jets matched to a generated top quark as a function of the generated top p_T, (b) t-mistag rate in the μ + jets channel for AK8 jets unmatched to a generated top quark as a function of the jet p_T.

Section 5.7; this is done because an independent measurement (performed in a sample exclusive to the SR of this analysis) of the data/MC SF for this t-tagging selection is currently not available.

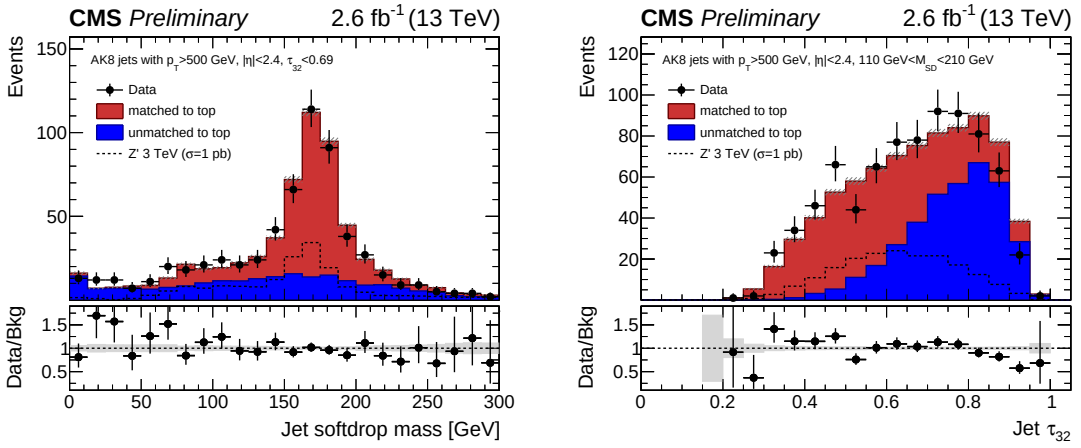
- the t-mistag efficiency is measured in data and simulation using two control samples (exclusive to the ℓ + jets SR) dominated by W + jets production; this provides a direct measurement of the data/MC SFs for the t-mistag rate, which is used to correct the simulation in the ℓ + jets SR. This study is presented in Section 5.4.3.1.

We check the modeling of the softdrop mass and τ_{32} in the simulation by comparing data and MC in a ℓ + jets event sample with slightly looser cuts compared to the ℓ + jets SR defined in Section 5.3.2. The only differences with respect to the standard ℓ + jets selection are given by the removal of the $H_{\mathrm{T}}^{\mathrm{lep}} > 150$ GeV cut in the μ + jets channel and and the reduction of the $\not\!\!E_T$ threshold from 120 GeV to 50 GeV in the e + jets channel. The softdrop mass and τ_{32} distributions for this looser ℓ + jets selection are displayed in Figure 5.29. Good agreement between data and simulation is observed for both variables. The same level of agreement is found when considering separately jets in the central ($|\eta^j| < 1.0$) and forward regions ($1.0 < |\eta^j| < 2.4$); this justifies the use of data/MC t-tagging SF not dependent on the jet pseudorapidity.

5.4.3.1 t-mistag rate measurement in data (W+jets CR)

We measure the t-mistag efficiency directly in data using a control sample dominated by W + jets production, which is the main non-$t\bar{t}$ background in the ℓ + jets analysis.

The mistag rate of the t-tagging selection is measured in two separate control regions, one for the μ + jets channel and one for e + jets channel. These two samples are defined as follows. Events in the μ + jets (e + jets) CR are required to pass the μ + jets (e + jets) preselection cuts listed in Section 5.3.2, except for a looser requirement on $H_{\mathrm{T}}^{\mathrm{lep}}$ ($\not\!\!E_T$): in



(a) (b)

Figure 5.29: Data/MC comparison for the two t-tagging variables used in the ℓ+jets analysis: (a) softdrop mass after cutting on τ_{32}, (b) τ_{32} after cutting on softdrop mass. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The total background is normalized to the data and its error band includes only the MC statistical uncertainty.

order to increase the statistics in these samples, the cut on $H_{\mathrm{T}}^{\mathrm{lep}}$ is removed in the $\mu+$jets CR and the $\not\!\!E_T$ threshold is lowered from 120 GeV to 50 GeV for the $e+$jets CR. In each CR, we require the presence of at least one AK8 jet candidate ($p_T > 500$ GeV, $|\eta| < 2.4$). Events with one or more AK4 jets passing a loose b-tagging working point are vetoed, in order to reduce the contribution of the $t\bar{t}$ background in these control samples. Finally, the orthogonality with the $\ell+$jets SR is ensured by inverting the cut on the leptonic term of χ^2 discriminator of the reconstructed $t\bar{t}$ system, given by $\chi^2_{\mathrm{lep}} > 30$.

A data/MC comparison for the kinematics of the AK8 jet candidates in the $\mu+$jets CR, before and after the t-tagging requirements, is given in Figure 5.30; similar distributions are obtained in the case of the $e+$jets CR. Both control samples are fully dominated by W+jets production, with a small contribution coming from SM $t\bar{t}$. On the other hand, the fraction of $t\bar{t}$ events increases considerably after the t-tagging requirements are applied; for this reason, when measuring the t-mistag efficiency in data, the yield expected from $t\bar{t}$ production is subtracted from data. The t-mistag rate measured in data thus corresponds to

$$\varepsilon^{\mathrm{mistag}}_{\mathrm{DATA}} = \frac{N^{\mathrm{tagged}}_{\mathrm{DATA}} - N^{\mathrm{tagged}}_{t\bar{t}}}{N_{\mathrm{DATA}} - N_{t\bar{t}}} \tag{5.6}$$

where N_{DATA} ($N_{t\bar{t}}$) denotes the number of AK8 jet candidates in the control sample for data ($t\bar{t}$ MC) and $N^{\mathrm{tagged}}_{\mathrm{DATA}}$ ($N^{\mathrm{tagged}}_{t\bar{t}}$) corresponds to the number of AK8 jets that are t-tagged in data ($t\bar{t}$ MC). For simulated events, the t-mistag rate is determined from the W + jets MC.

Table 5.5 shows the t-mistag rate measured in data and MC, together with the corresponding data/MC SF; separate values are shown for the $\mu+$jets and $e+$jets CRs. The SFs measured in the two CRs are found to be compatible. The combined data/MC SF corresponds to $\mathrm{SF}_\ell = 0.82 \pm 0.17$. This scale factor is used in the $\ell+$jets SR to correct the mis-identification efficiency of the t-tagging algorithm.

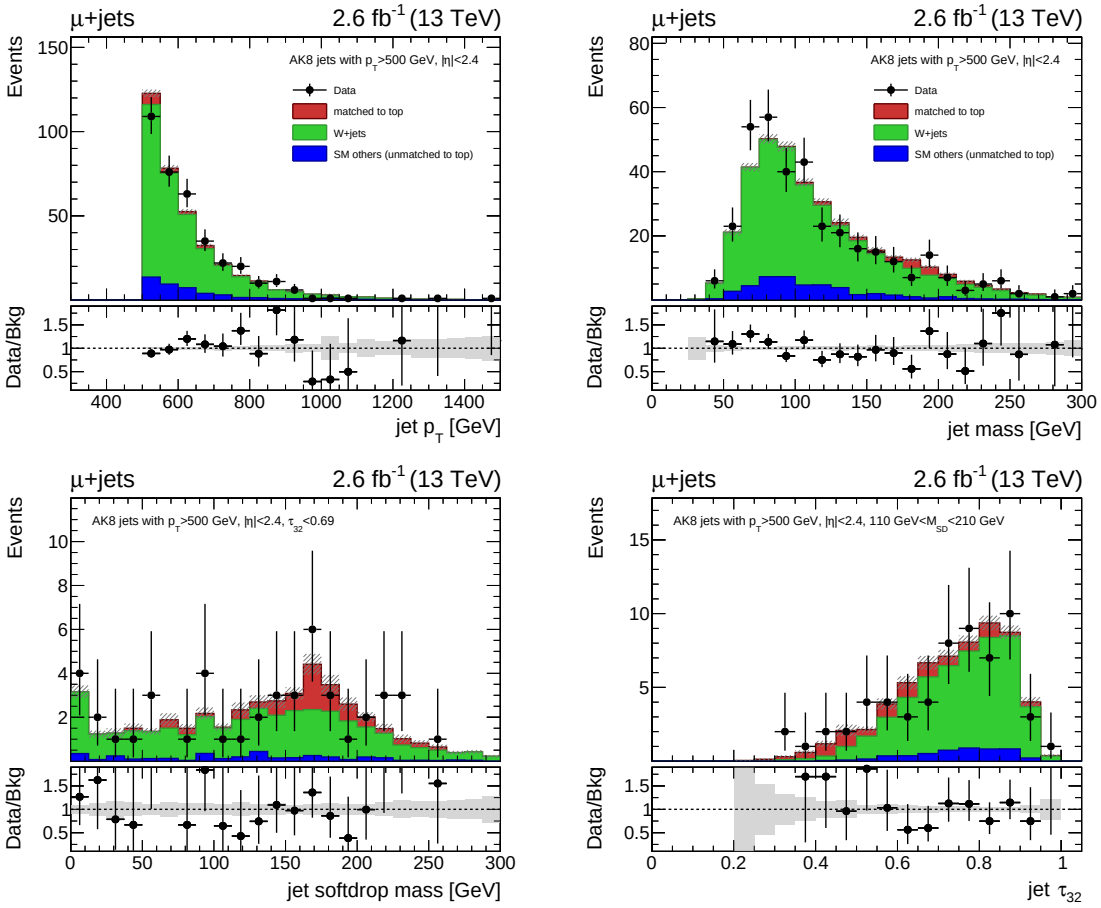


Figure 5.30: Data/MC distributions of AK8 jets in the W($\to \mu\nu$) + jets CR for the t-mistag rate measurement: (top) p_T and mass distributions for all jet candidates before any t-tagging requirement, (bottom) jet softdrop mass after cutting on the τ_{32} ratio and vice versa. In each plot, the total expected background is normalized to the data and the background error band includes only the MC statistical uncertainty.

Table 5.5: t-mistag efficiencies on AK8 jets for data and MC, plus their ratio, in the W + jets control regions.

t-mistag CR	$\varepsilon_{\text{Data}}$	ε_{MC}	SF
μ + jets	0.042 ± 0.012	0.050 ± 0.002	0.83 ± 0.24
e + jets	0.038 ± 0.010	0.046 ± 0.007	0.82 ± 0.25

5.5 Systematic uncertainties

This section details the systematic uncertainties considered in the ℓ + jets analysis for the MC predictions of background and signal processes. These uncertainties can affect both the normalization and shape of the distributions of physics observables, most notably the invariant mass of the reconstructed $t\bar{t}$ system.

The impact of each systematic uncertainty on the normalization of the $M_{t\bar{t}}$ distribution for background and signal processes is given in Table 5.6. The uncertainties on the cross sections of SM backgrounds and on the total integrated luminosity only enter as normalization uncertainties for each given process; all the other systematic uncertainties can affect both the normalization and shape of the $M_{t\bar{t}}$ distributions of each of the processes they apply to. The main systematic errors in this analysis are due to the data/MC corrections for the jet b-tagging and t-tagging efficiencies, the uncertainties on the normalization of the SM backgrounds and the theoretical uncertainties related to the choice of the PDF set and the renormalization and factorization scales used in the simulation of background and signal processes.

Each of the systematic uncertainties considered in the analysis is described below.

Standard Model cross sections

The uncertainties on the cross sections for $t\bar{t}$, W + jets and Z + jets production are derived from the background fit described in Section 5.6: they correspond, respectively, to 8% ($t\bar{t}$), 6% (W + jets) and 20% (Z + jets). An uncertainty of 20% is assigned to the cross section of single-top and diboson production; the latter processes, together with Z + jets production, are expected to be small backgrounds in the ℓ + jets analysis.

Integrated luminosity

The uncertainty on the total integrated luminosity recorded by CMS in the 2015 data-taking period at $\sqrt{s} = 13$ TeV is 2.7% [147].

Pileup reweighting

The systematic uncertainty associated to the pileup correction is evaluated by varying the value of the minimum-bias cross section by ±5%. The corresponding uncertainty on the acceptance of signal and background events is very small, below 1%.

Muon ID and HLT efficiencies

The uncertainties on the muon ID and HLT efficiencies are determined by varying the corresponding data/MC SFs with their $\pm 1\sigma$ errors, as a function of the muon p_T and η. The systematic error due to the muon ID (HLT) efficiency corrections corresponds to approximately a 2% (1%) uncertainty.

Electron ID and HLT efficiencies

The uncertainties on the data/MC SFs used for the electron ID efficiency are propagated to obtain the corresponding systematic error; this amounts to approximately a 1% uncertainty. A systematic uncertainty of 2% is assigned to the data/MC SF for the efficiency of the e+jets trigger, based on the measurement described in Section 5.4.2.

Jet energy scale (JES)

The systematic uncertainty due to the JEC is estimated by changing the jet energy scale corrections in MC within their $\pm 1\sigma$ errors. This systematic error affects the acceptance of background and signal processes up to 4%; more importantly, it can change significantly the shape of the $M_{t\bar{t}}$ distributions. Figure 5.31 shows the impact of the JEC variations on the $t\bar{t}$ background and a high-mass signal sample. The uncertainties on the JEC for AK4 and AK8 jets are treated as fully correlated and their effect is propagated to the $\not{E}_T$ measurement.

Jet energy resolution (JER)

As for the JEC systematic error, the uncertainties on the JER correction factors are propagated to the final MC predictions. The resulting error is at the level of 1% and it is thus smaller compared to the JEC uncertainty. Similarly to the uncertainties for the energy scale corrections, the JER uncertainties for AK4 and AK8 jets are treated as fully correlated and their effect is propagated to the $\not{E}_T$ measurement.

b-tagging efficiency and mistag rate

The b-tagging efficiency SFs [86], used in the reweighting procedure detailed in Section 5.4.3, are varied within their $\pm 1\sigma$ errors in order to determine the corresponding uncertainty on the final MC predictions. The uncertainties of SF_b and SF_c are treated as fully correlated, whereas the uncertainty on SF_l (b-mistag rate) is treated as uncorrelated from those on SF_b and SF_c. As shown in Table 5.6, the size of each of these uncertainties strongly depends on the process and event category considered. The systematic variations due to the b-tagging SF uncertainties for the W + jets, $t\bar{t}$ and Z′ samples are displayed in Figure 5.32.

t-tagging efficiency and mistag rate

The systematic uncertainty related to the t-tagging efficiency SFs are determined using the same method described above for the b-tagging uncertainties. As described in Section 5.4.3.1, the data/MC SF for the t-mistag rate is measured in a control region with a 21% uncertainty ($\mathrm{SF}_l = 0.82 \pm 0.17$). Due to the lack of an independent measurement of the scale factor for the t-tagging efficiency, this uncertainty is ultimately introduced in the statistical analysis as an unconstrained (free) nuisance parameter. The uncertainties of SF_t and SF_l are treated as uncorrelated.

Parton Distribution Functions (PDFs)

All simulated samples for background and signal processes are generated using PDFs from the NNPDF-3.0 set [137]. The corresponding systematic uncertainty on the $M_{t\bar{t}}$ distributions is determined according to the procedure described in Ref. [130], for both background and signal processes.

ME Q^2-scale uncertainty for $t\bar{t}$ and W + jets production

The effect due to missing higher orders in the simulation of SM $t\bar{t}$ and W + jets production is estimated by varying up and down the renormalization and factorization scales used in the matrix element of these MC simulations. The effect of this systematic uncertainty can be seen in Figure 5.33. The scale uncertainties in the $t\bar{t}$ MC and in the W + jets MC are considered uncorrelated.

Parton shower scale for $t\bar{t}$ production

We assign an additional uncertainty to the $t\bar{t}$ MC, related to the modeling of the parton shower in the simulation. This uncertainty is determined using dedicated samples, in which the parton shower p_T^2-scale is varied by a factor of 1/4 (down) and 4 (up). Figure 5.33 shows the effect of this systematic uncertainty on a background and a signal sample.

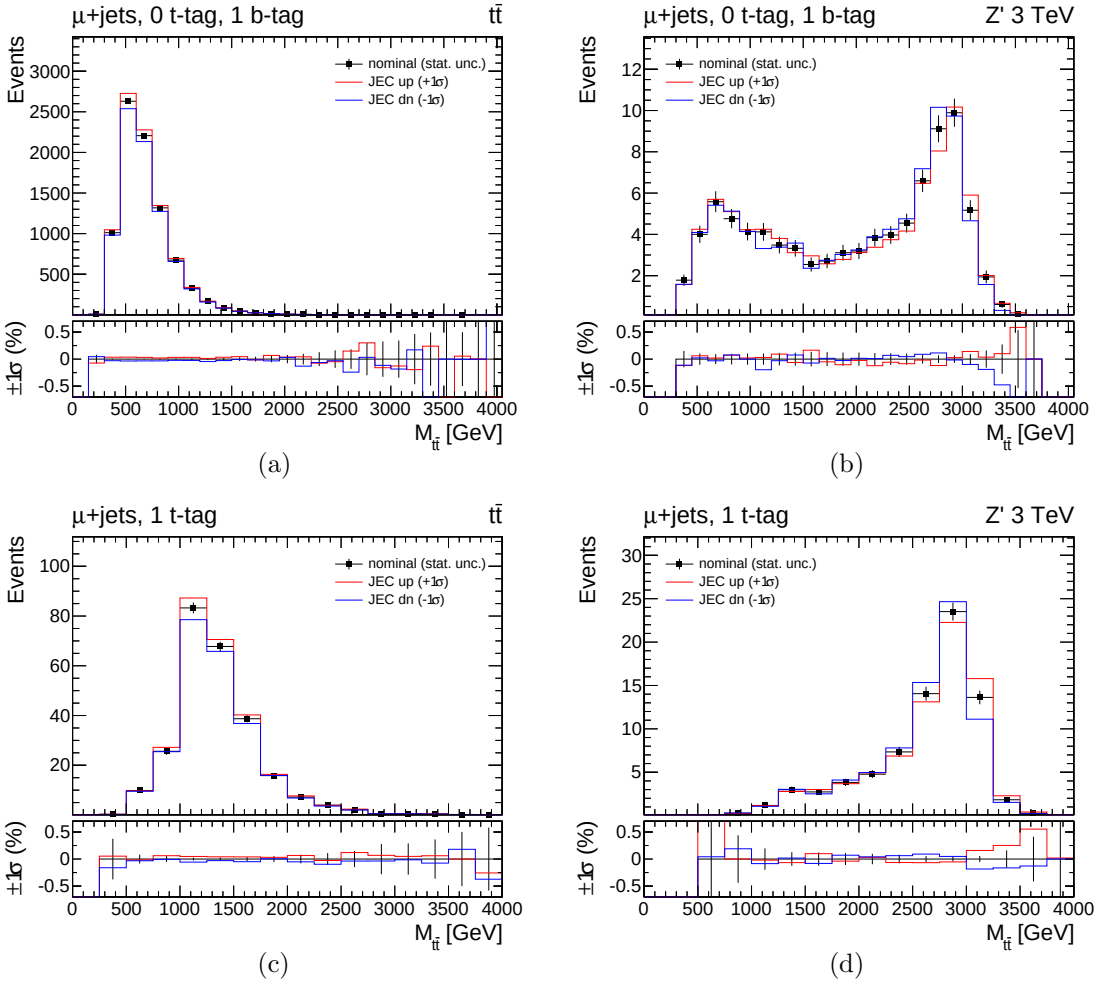


Figure 5.31: Variation of the $M_{t\bar{t}}$ distribution due to the systematic uncertainty on the jet energy corrections: (a, c) $t\bar{t}$ background, (b, d) Z′ signal with $M_{Z'} = 3$ TeV. The distributions on the top (bottom) refer to the `T0B1` (`T1`) category of the μ + jets analysis.

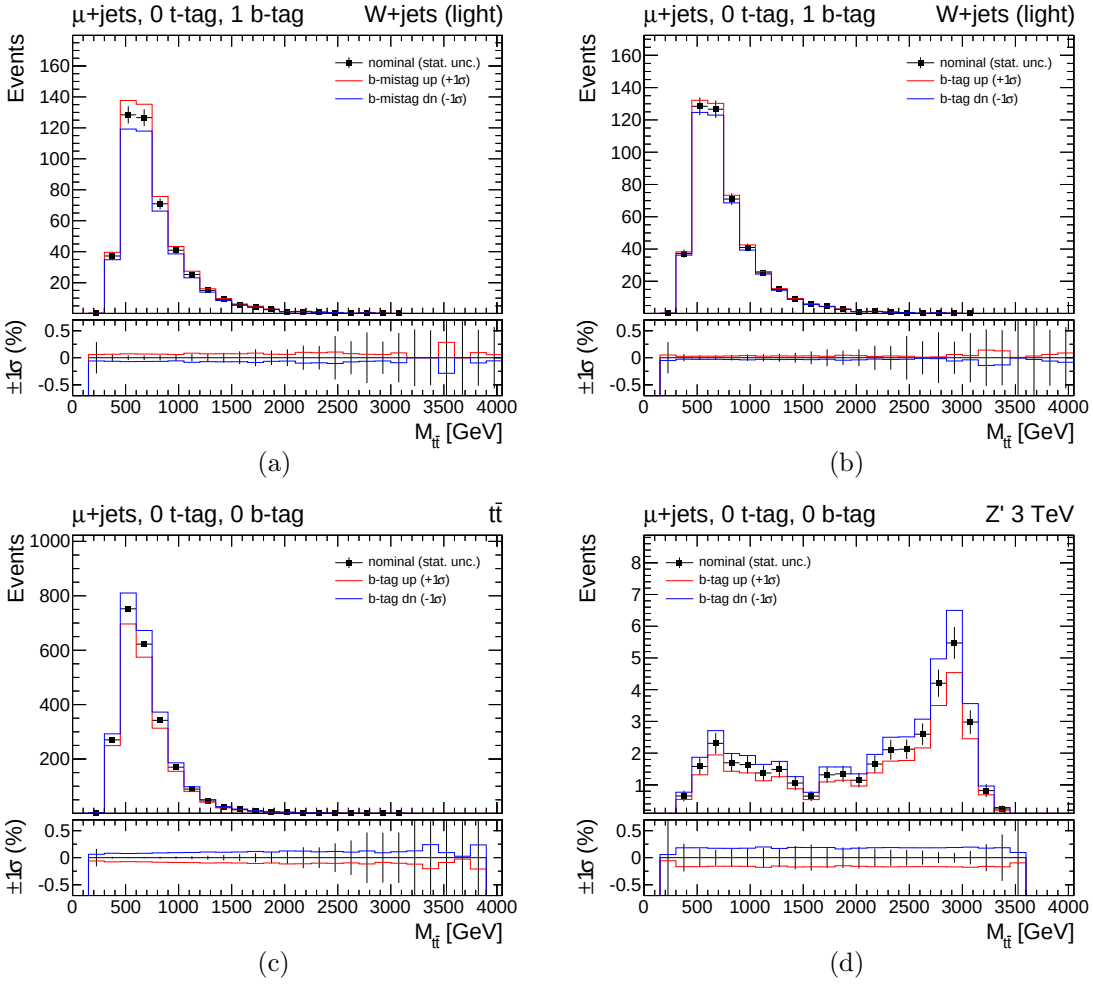


Figure 5.32: Variation of the $M_{t\bar{t}}$ distribution due to the systematic uncertainty on the b-tagging efficiency and mistag rate. Top: b-mistag rate (a) and b-tagging efficiency (b) systematics for the W + jets MC. Bottom: systematic uncertainty due to the b-tagging efficiency for the $t\bar{t}$ background (c) and a Z′ signal with $M_{Z'} = 3$ TeV (d) in the `TOBO` category of the μ + jets analysis.

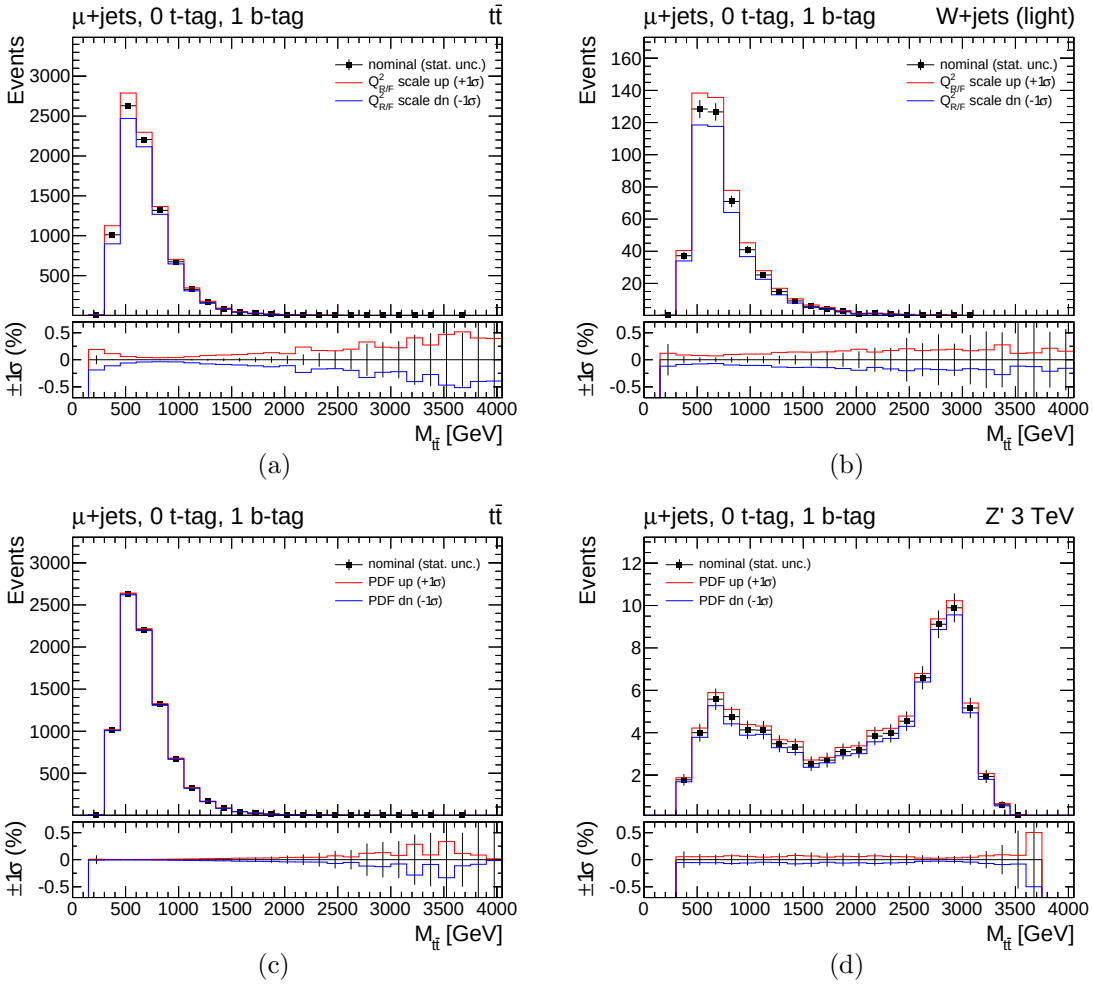


Figure 5.33: Variation of the $M_{t\bar{t}}$ distribution due to the theoretical uncertainties associated to the MC simulation: Top: ME Q^2-scale uncertainties for the $t\bar{t}$ (a) and W+jets (b) MC samples. Bottom: PDF uncertainty for $t\bar{t}$ background (c) and a Z′ signal with $M_{Z'} = 3$ TeV (d). Each of these distributions refers to the `T0B1` category of the μ + jets analysis.

Table 5.6: Summary of the impact of the systematic uncertainties on the normalization of background and signal processes. The values in the table correspond to the percentage difference with respect to the nominal yield. When necessary, the effect of a given uncertainty is reported separately for different event categories. Squared brackets, when present, denote the range of a given uncertainty across different categories or processes. The +/− symbols, when present, indicate the correlation between the systematic uncertainty and the variation of the MC yield. For the t-tagging efficiency SF, results are given for an arbitrary uncertainty of 25%, used only for illustration purposes.

Systematic uncertainty	Background	Z′ signal (1% width)		
		$M = 2$ TeV	$M = 3$ TeV	$M = 4$ TeV
SM cross sections	8% ($t\bar{t}$) 6% (W + jets) 20% (Z + jets) 20% (single-top) 20% (diboson)			
luminosity	+2.7%			
pileup	0.6%	0.5%	1.1%	0.7%
electron ID	+0.7%	+0.8%	+0.9%	+1.0%
muon ID	+1.6%	+1.6%	+1.8%	+1.9%
electron trigger	+2%			
muon trigger	+0.8%	+0.9%	+1.0%	+1.0%
JEC	[+3.0%, +4.6%]	0.8%		
JER	0.5%	−0.8%		
b-mistag rate (`T0B0`)	−0.8%	−0.5%		
b-mistag rate (`T0B1`)	+0.2% ($t\bar{t}$) , +5.0% (W + jets)	+0.5%	+0.8%	+1.0%
b-mistag rate (`T1`)	+0.2%	+0.3%	+0.4%	+0.3%
b-tagging efficiency (`T0B0`)	−8.1% ($t\bar{t}$) , −0.7% (W + jets)	−14%	−17%	−18%
b-tagging efficiency (`T0B1`)	+2% ($t\bar{t}$) , +4% (W + jets)	+3%		
b-tagging efficiency (`T1`)	+0.2%	−0.2%	−3%	−5%
t-mistag rate (`T0B0`, `T0B1`)	−0.2%	−0.1%		
t-mistag rate (`T1`)	+1.2% ($t\bar{t}$) , +21% (W + jets)	+0.1%	+0.3%	+0.4%
t-tagging efficiency (`T0B0`, `T0B1`)	−2%	−22%	−19%	−18%
t-tagging efficiency (`T1`)	+23%	+25%	+24%	+24%
parton shower scale ($t\bar{t}$)	7%			
ME Q^2-scales, $t\bar{t}$ (`T0B0`, `T0B1`)	[7%, 15%]			
ME Q^2-scales, $t\bar{t}$ (`T1`)	6%			
ME Q^2-scales, W + jets (`T0B0`, `T0B1`)	[8%, 11%]			
ME Q^2-scales, W + jets (`T1`)	15%			
PDFs	[2%, 4%]	3%	6%	12%

5.6 Background estimation

Several SM processes contribute to the selected event sample in different proportions, depending on the event category considered. The two main backgrounds in the analysis are SM $t\bar{t}$ and W + jets production. The latter accounts for a sizeable portion of the total background in the `T0B0` category, whereas the former fully dominates the `T0B1` and `T1` categories. Single-top, Z + jets and diboson production are overall small backgrounds in all the ℓ + jets SR categories. As already mentioned in previous sections, the distributions obtained from MC simulations are properly reweighted to account for known discrepancies between data and simulated events; these include the pileup multiplicity, the trigger efficiencies, the lepton-ID efficiencies and the efficiencies for jet b-tagging and t-tagging. A detailed description on these corrections and the corresponding systematic uncertainties can be found in Section 5.5.

The final background estimates in this search are determined by fitting the background-only hypothesis to the data. The fit is performed using a binned maximum-likelihood method, as implemented in the THETA package [131]. Four parameters are left unconstrained in the fitting procedure: these are the parameters associated to cross sections for $t\bar{t}$, W + jets and Z + jets, and the data/MC SF for the t-tagging efficiency. Leaving these four parameters free to float in the fit improves the flexibility of the background model; moreover, it allows to estimate their uncertainties in the phase space considered by this search with a data-driven procedure. All the remaining systematic uncertainties, described in Section 5.5, are included in the fit as nuisance parameters. Nuisance parameters associated to normalization uncertainties are assigned a log-normal prior, while a gaussian prior distribution is used for each of the systematic uncertainties that affect the shape of the $M_{t\bar{t}}$ distributions.

The SM backgrounds considered in the fit are $t\bar{t}$, W + jets, single-top, Z + jets and diboson production. The prediction of the W + jets simulation is split into two separate processes, W + light partons and W + heavy-flavor quarks, based on the type of partons generated in the MC matrix element. The two processes are assigned the same systematic uncertainties, but the effect of each systematic can be different in the two cases, so the proportion of light- and heavy-flavor W + jets events can be varied by the likelihood fit. Observables in different event categories are fitted simultaneously to find the best-fit value of the background prediction, correcting the central values of the background nuisance parameters. The following observables are used in the fit:

- low-$M_{t\bar{t}}$ SR observables: reconstructed $M_{t\bar{t}}$ distributions in the low-$M_{t\bar{t}}$ ℓ + jets SR ($M_{t\bar{t}} < 2$ TeV, $\chi^2 < 30$);
- high-χ^2 CR observables: reconstructed $M_{t\bar{t}}$ distributions in the high-χ^2 ℓ + jets CR, given by $\chi^2 > 30$;
- dimuon CR observable: dimuon invariant mass in a $\mu^+\mu^-$ CR obtained by removing only the (same-flavor) lepton veto from the ℓ + jets selection and adding a Z-mass window requirement (71 GeV $< M_{\ell^+\ell^-} <$ 111 GeV).

In the case of the low-$M_{t\bar{t}}$ ℓ + jets SR and the high-χ^2 ℓ + jets CR, events are also split in six exclusive categories, based on lepton flavor and number of b-tagged and t-tagged jets (`T1`, `T0B1`, `T0B0`). Based on the above, a total of thirteen observables are used in the fit. The high-χ^2 CR is dominated by W + jets production, which is also the main non-$t\bar{t}$ background in the ℓ + jets SR. The dimuon CR is sensitive to Drell-Yan production and it is used to constrain the uncertainty on the cross section of this SM background. It is worth noting that this dilepton CR includes cuts on the missing E_T ($\not{E}_T > 50$ GeV) and the χ^2 discriminator

($\chi^2 < 30$), in order to be as similar as possible to the ℓ + jets selection. The corresponding dielectron control region is not used because of its limited statistics, mainly due to the tighter threshold on the missing transverse energy ($\not{E}_T > 120$ GeV).

The post-fit nuisance parameters, expressed in units of their prior uncertainties, are shown in Figure 5.34. Figure 5.34(a) shows the results of the ML fit described above, combining SR and CR categories; Figure 5.34(b) and Figure 5.34(c) include the results obtained from the fit using, respectively, only the six low-$M_{t\bar{t}}$ SR observables and only the seven CR observables. The latter two fits are only used to cross-check the results of the combined fit in Figure 5.34(a). No post-fit parameter lies outside of 2σ of its prior uncertainty. The results of the main fit are consistent with those obtained with the two complementary fits shown in Figure 5.34(b) and Figure 5.34(c). By comparing the three cases, it is possible to see how different parameters are constrained by certain observables: in the case of the CR-only fit, there is limited sensitivity to the t-tagging efficiency, compared to the SR-only fit, but the parameters related to the modeling of the V + jets background are better constrained. As expected, the combined fit provides the best constraining power for each of the fit parameters. The uncertainties obtained for the four free parameters of the fit are 8% for the $t\bar{t}$ cross section, 6% for the W + jets cross section, 20% for the Z + jets cross section and, finally, 6% for the t-tagging efficiency SF.

The event yields observed in data and expected from SM processes after the background fit are given in Table 5.7 for the V + jets CRs and in Table 5.8 for the six SR categories. The corresponding distributions employed in the fit are plotted in Figures 5.35–5.39. In addition, Figures 5.40–5.46 include a set of data/MC comparisons for basic kinematic quantities used in the ℓ + jets selection.

Good agreement between data and background prediction is observed for all the distributions considered. Residual differences between data and expected SM yields are covered by the systematic uncertainties described in Section 5.5. Based on these results, no significant deviation from the background expected from SM processes is observed in the data.

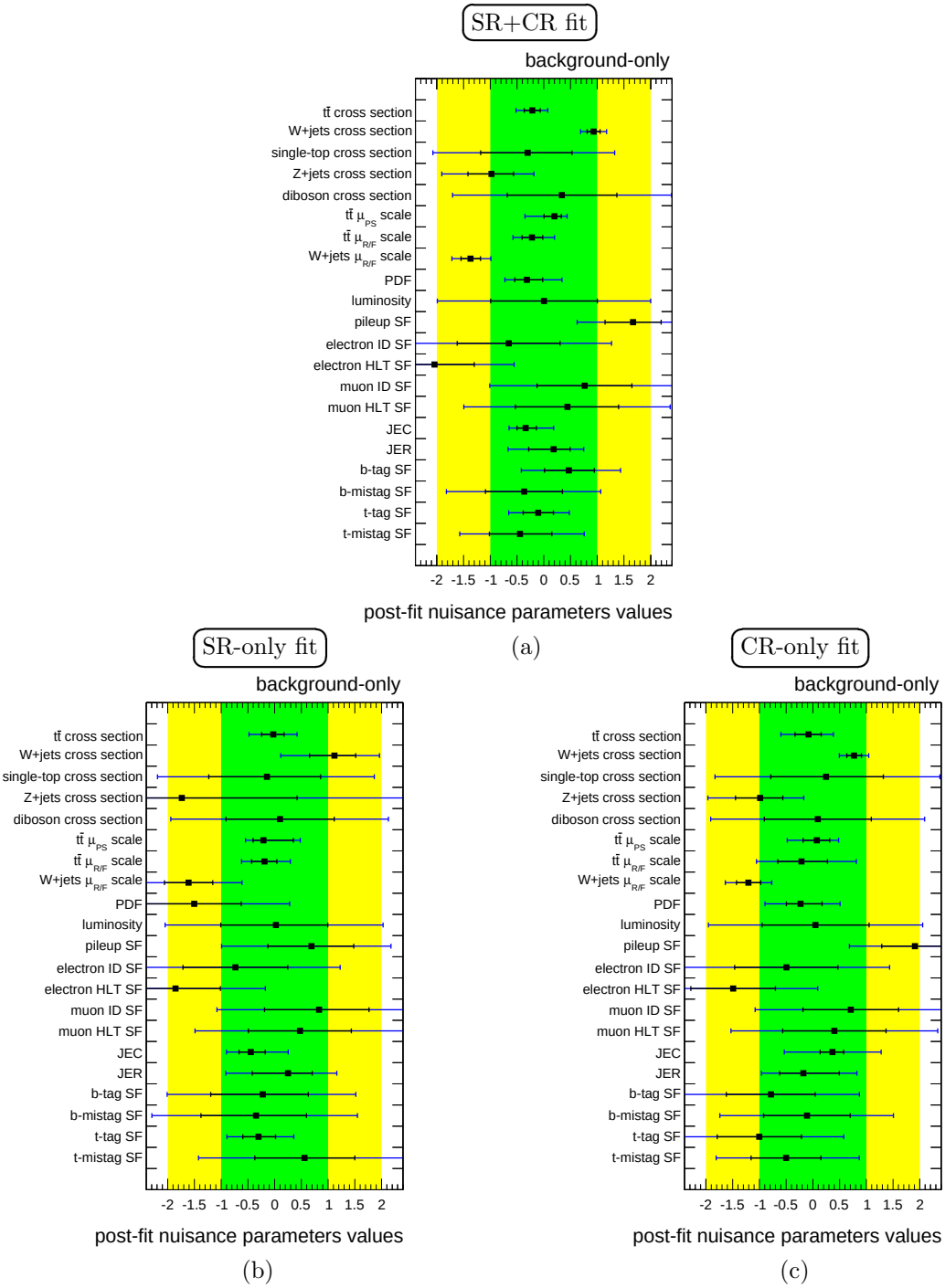


Figure 5.34: Post-fit values of the nuisance parameters of the background model: (a) fit results obtained from the maximum-likelihood fit to data described in the text, using simultaneously the low-$M_{t\bar{t}}$ SR and the V + jets CR observables; (b,c) fit results using only the low-$M_{t\bar{t}}$ SR observables (b) and only the V + jets CR observables (c). The fit parameters are expressed as deviations from their pre-fit values (zero), in units of the corresponding prior uncertainty. The post-fit uncertainty on the t-tagging data/MC SF is normalized to a prior uncertainty of 25%; the prior uncertainty used as reference for the $t\bar{t}$, W + jets and Z + jets cross sections corresponds to 50%. The 1σ (2σ) band of each parameter is shown in black (blue).

Table 5.7: Event yields in the high-χ^2 and dilepton control regions dominated by W + jets and Z+jets production, respectively. The expected yields for SM backgrounds are obtained from the ML fit to the data described in the text. The error associated to the total expected background includes the MC statistical uncertainty and all the post-fit systematic uncertainties.

Process	μ + jets CR ($\chi^2 > 30$)			e + jets CR ($\chi^2 > 30$)			$\mu\mu$ + jets CR
	T1	T0B1	T0B0	T1	T0B1	T0B0	
$t\bar{t}$	22 ± 3	2886 ± 306	1279 ± 141	11 ± 2	339 ± 41	152 ± 19	36 ± 4
W + jets (light-f)	24 ± 3	896 ± 90	13376 ± 1055	13 ± 2	180 ± 19	2155 ± 176	0.2 ± 0.1
W + jets (heavy-f)	4 ± 1	646 ± 57	2024 ± 165	2 ± 0	92 ± 9	246 ± 21	0.6 ± 0.3
single-top + DY + VV	9 ± 2	759 ± 120	1317 ± 173	6 ± 1	132 ± 22	172 ± 22	62 ± 13
Total Background	59 ± 6	5186 ± 387	17996 ± 1276	31 ± 3	744 ± 58	2725 ± 203	99 ± 14
Data	46	5197	18017	39	722	2753	106

Table 5.8: Event yields in the six ℓ + jets SR categories. The expected yields for SM backgrounds are obtained from the ML fit to the data described in the text. The error associated to the total expected background includes the MC statistical uncertainty and all the post-fit systematic uncertainties.

Process	μ + jets SR ($\chi^2 < 30$)			e + jets SR ($\chi^2 < 30$)		
	T1	T0B1	T0B0	T1	T0B1	T0B0
$t\bar{t}$	218 ± 28	7602 ± 826	1965 ± 229	119 ± 15	1016 ± 124	248 ± 32
W + jets (light-f)	27 ± 4	547 ± 54	4675 ± 377	13 ± 2	97 ± 10	684 ± 58
W + jets (heavy-f)	4 ± 1	333 ± 30	780 ± 65	2 ± 0	44 ± 4	84 ± 8
single-top + DY + VV	9 ± 2	682 ± 111	635 ± 85	4 ± 1	103 ± 18	74 ± 10
Total Background	258 ± 29	9164 ± 856	8055 ± 541	138 ± 16	1260 ± 129	1090 ± 78
Data	252	9230	7966	142	1217	1005

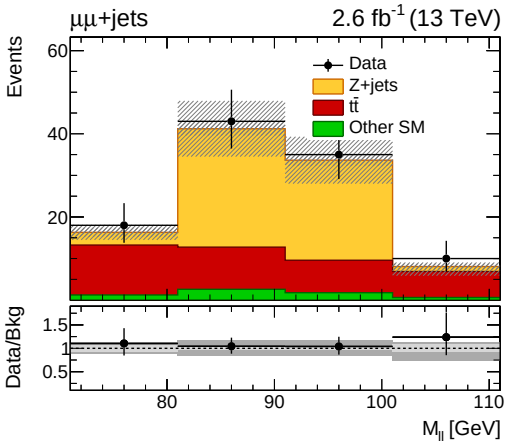


Figure 5.35: Dilepton invariant mass in the dimuon CR for data and expected background, after the background-only fit. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

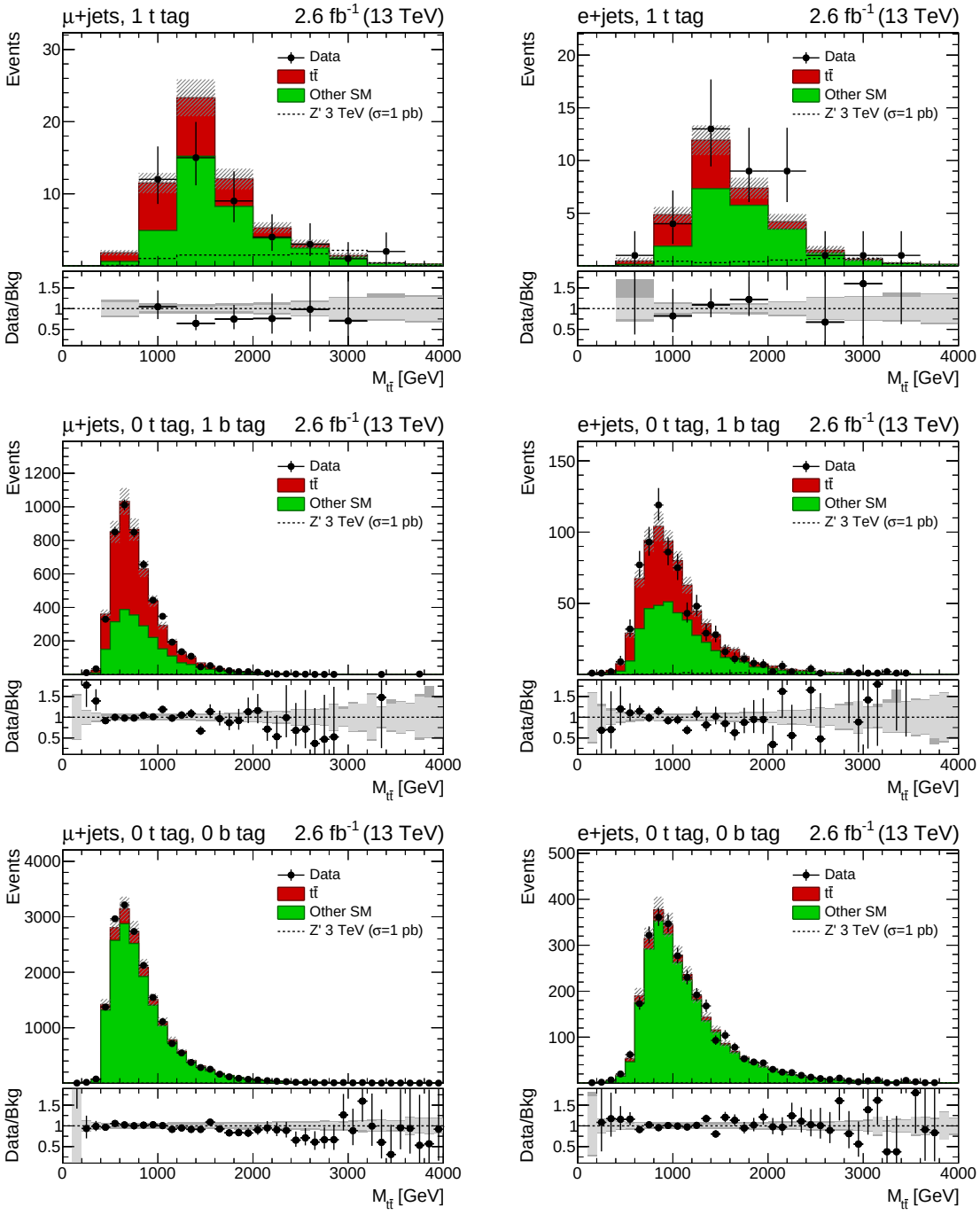


Figure 5.36: Invariant mass of the reconstructed $t\bar{t}$ pair for data and expected background in events passing the W + jets CR selection ($\chi^2 > 30$) after the background-only fit. Plots for the μ + jets (e + jets) channel are shown on the left (right). For each lepton flavor, events are split in three exclusive categories: (from top to bottom) `T1`, `T0B1` and `T0B0`. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

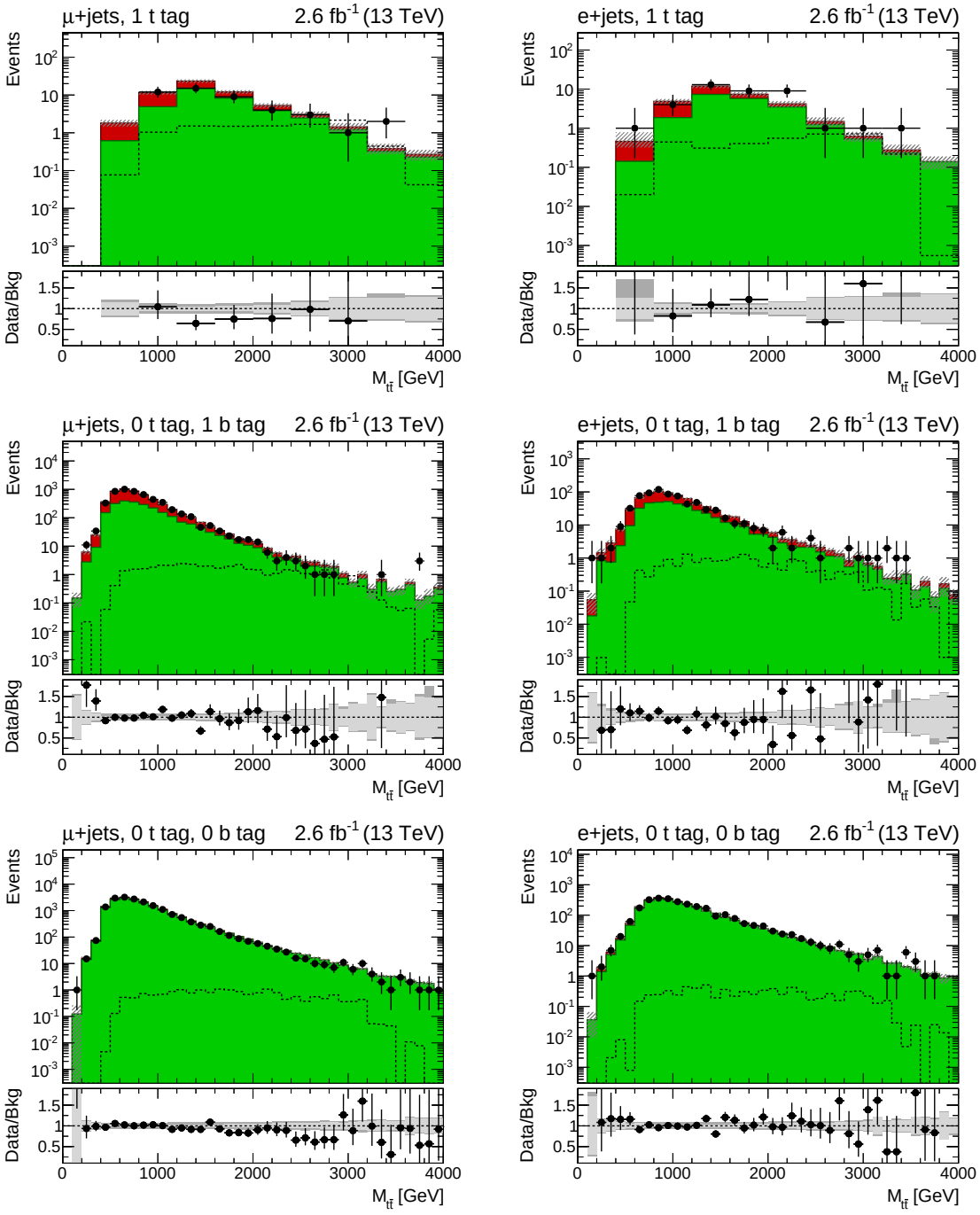


Figure 5.37: Distributions of Figure 5.36 in logarithmic scale. The legend of each plot is the same shown in Figure 5.36.

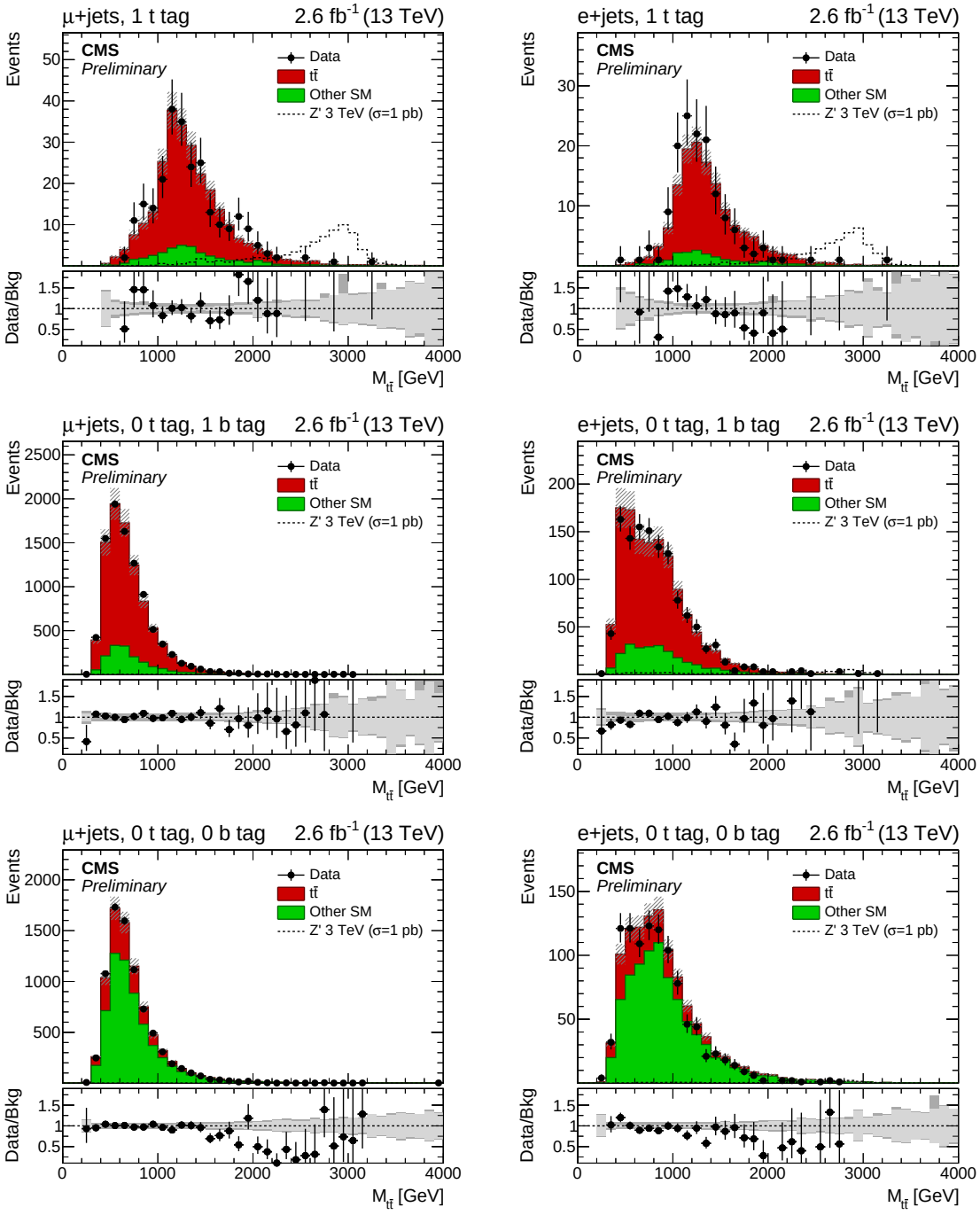


Figure 5.38: Invariant mass of the reconstructed $t\bar{t}$ pair for data and expected background in events passing the SR selection ($\chi^2 < 30$) after the background-only fit [4]. Plots for the μ + jets (e + jets) channel are shown on the left (right). For each lepton flavor, events are split in three exclusive categories: (from top to bottom) `T1`, `T0B1` and `T0B0`. Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

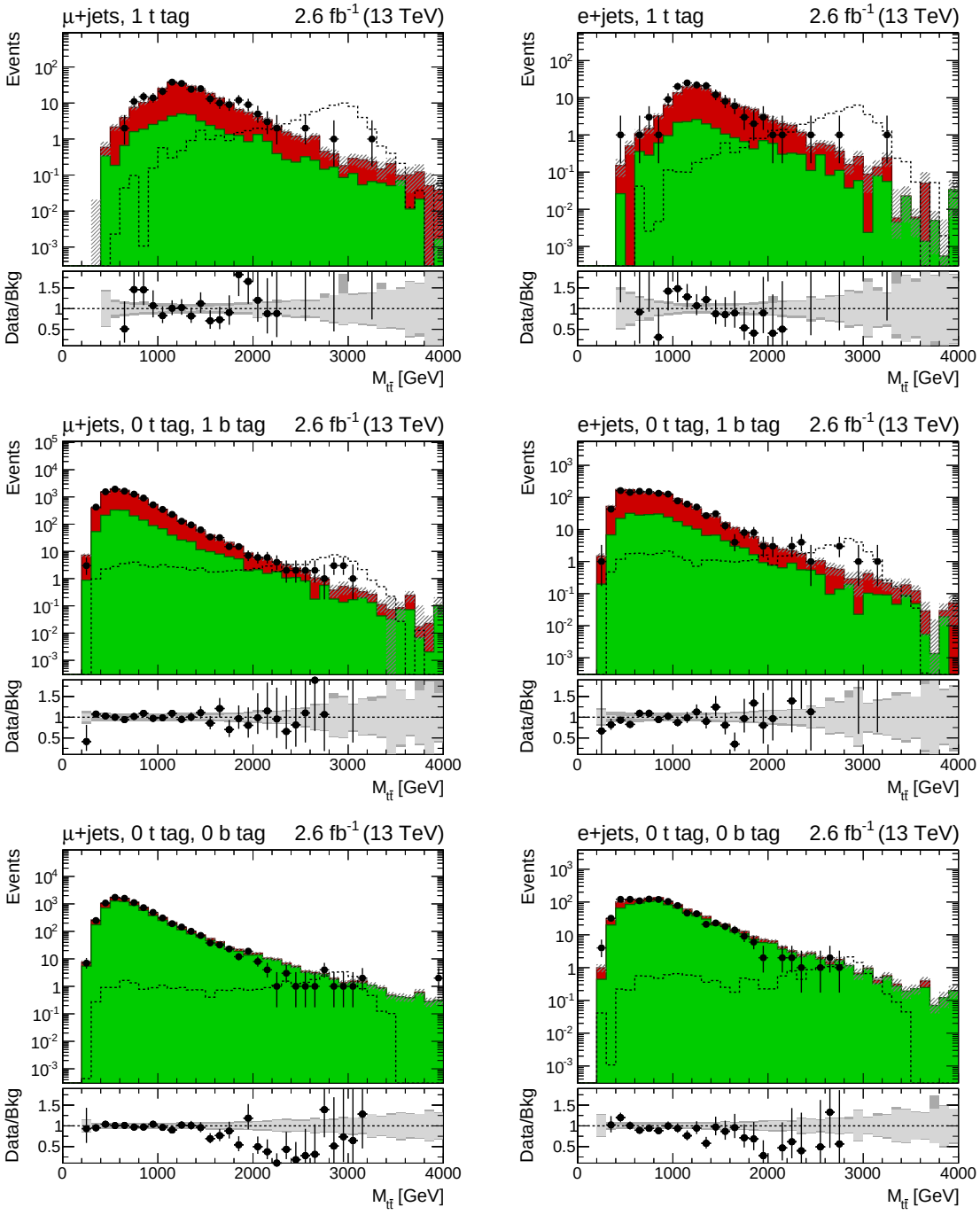


Figure 5.39: Distributions of Figure 5.38 in logarithmic scale. The legend of each plot is the same shown in Figure 5.38.

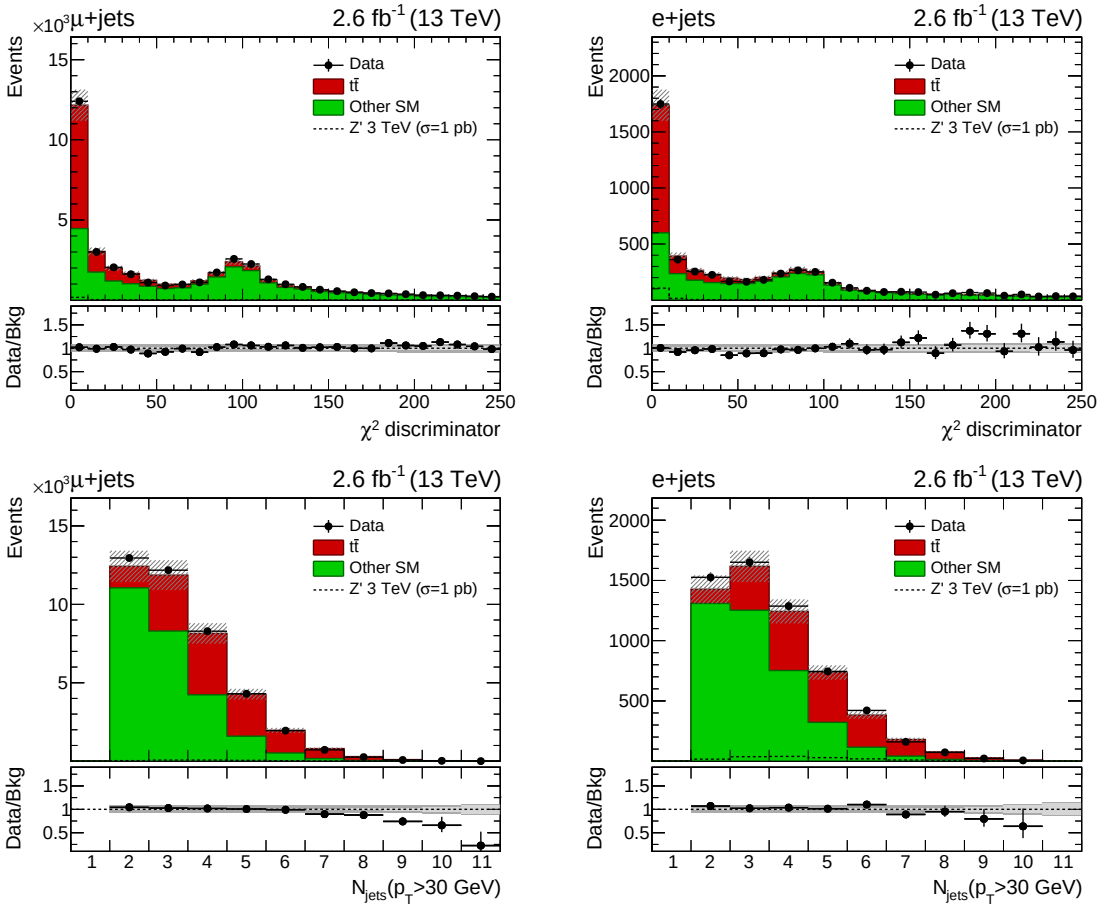


Figure 5.40: Data/MC comparison for the χ^2 discriminator of the reconstructed $t\bar{t}$ hypothesis and the number of reconstructed AK4 jets in events passing the ℓ + jets preselection (before χ^2 cut). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

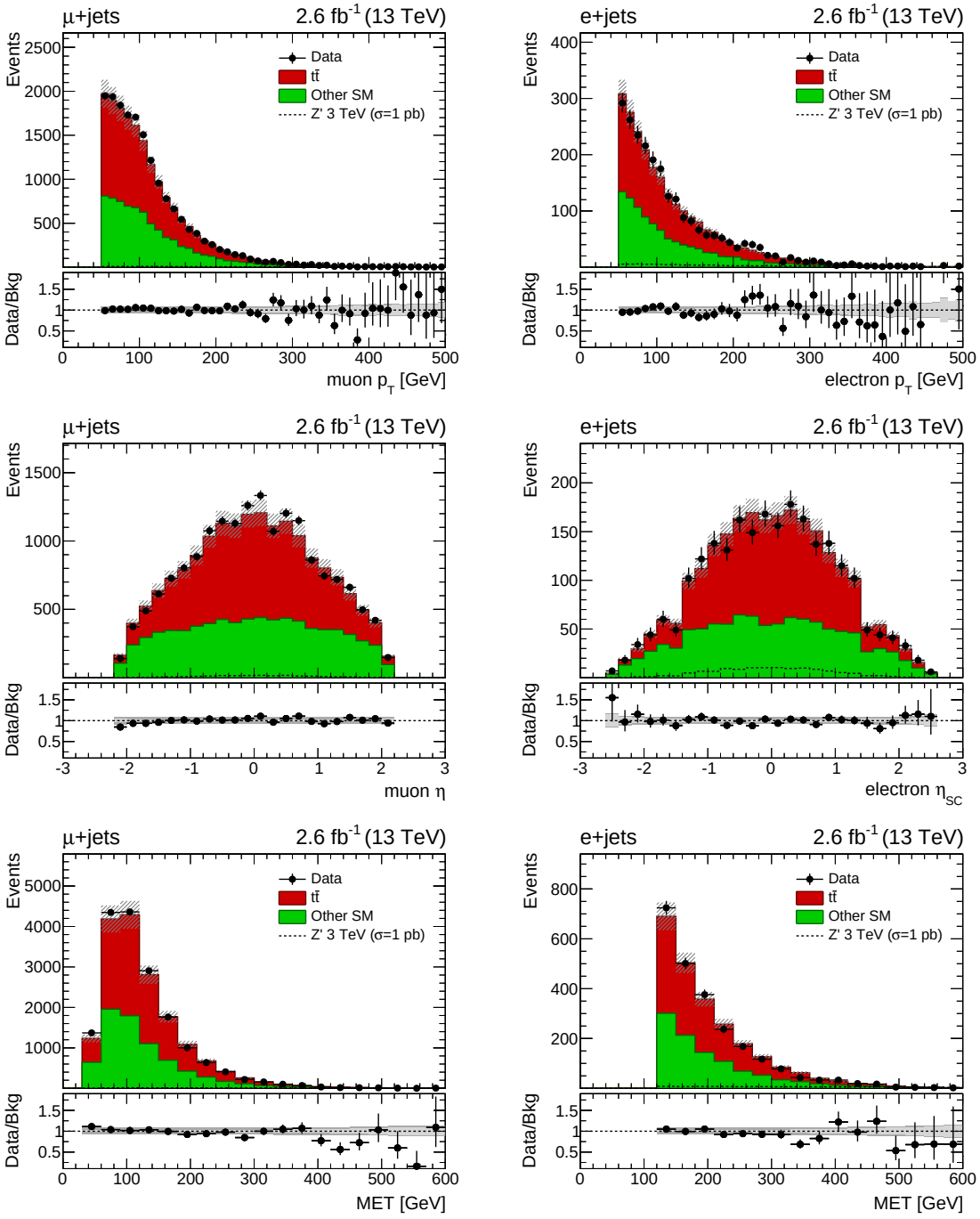


Figure 5.41: Data/MC comparison for the lepton p_T, the lepton η and the missing transverse energy in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

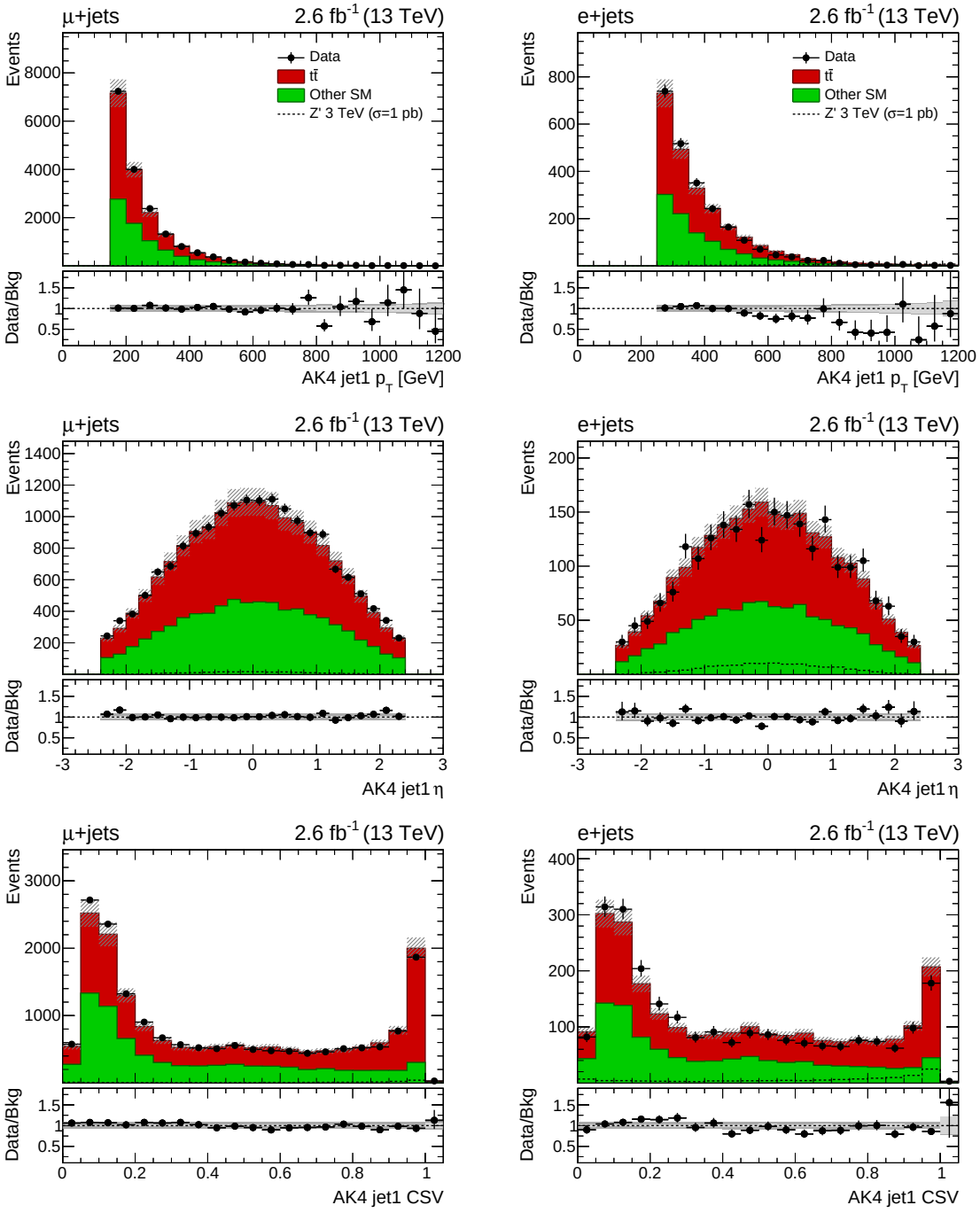


Figure 5.42: Data/MC comparison for the leading jet p_T, η and CSV score in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

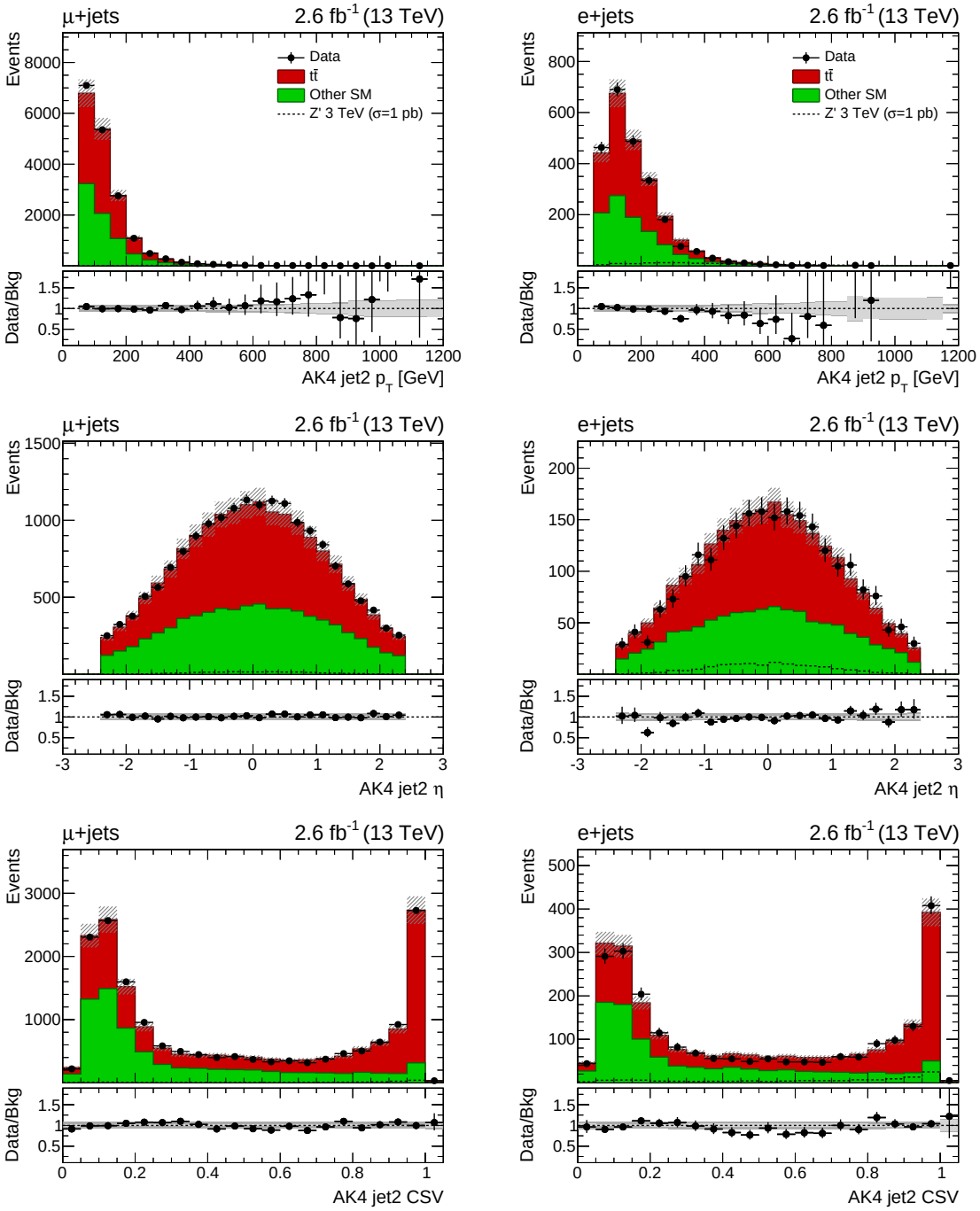


Figure 5.43: Data/MC comparison for the subleading jet p_T, η and CSV score in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

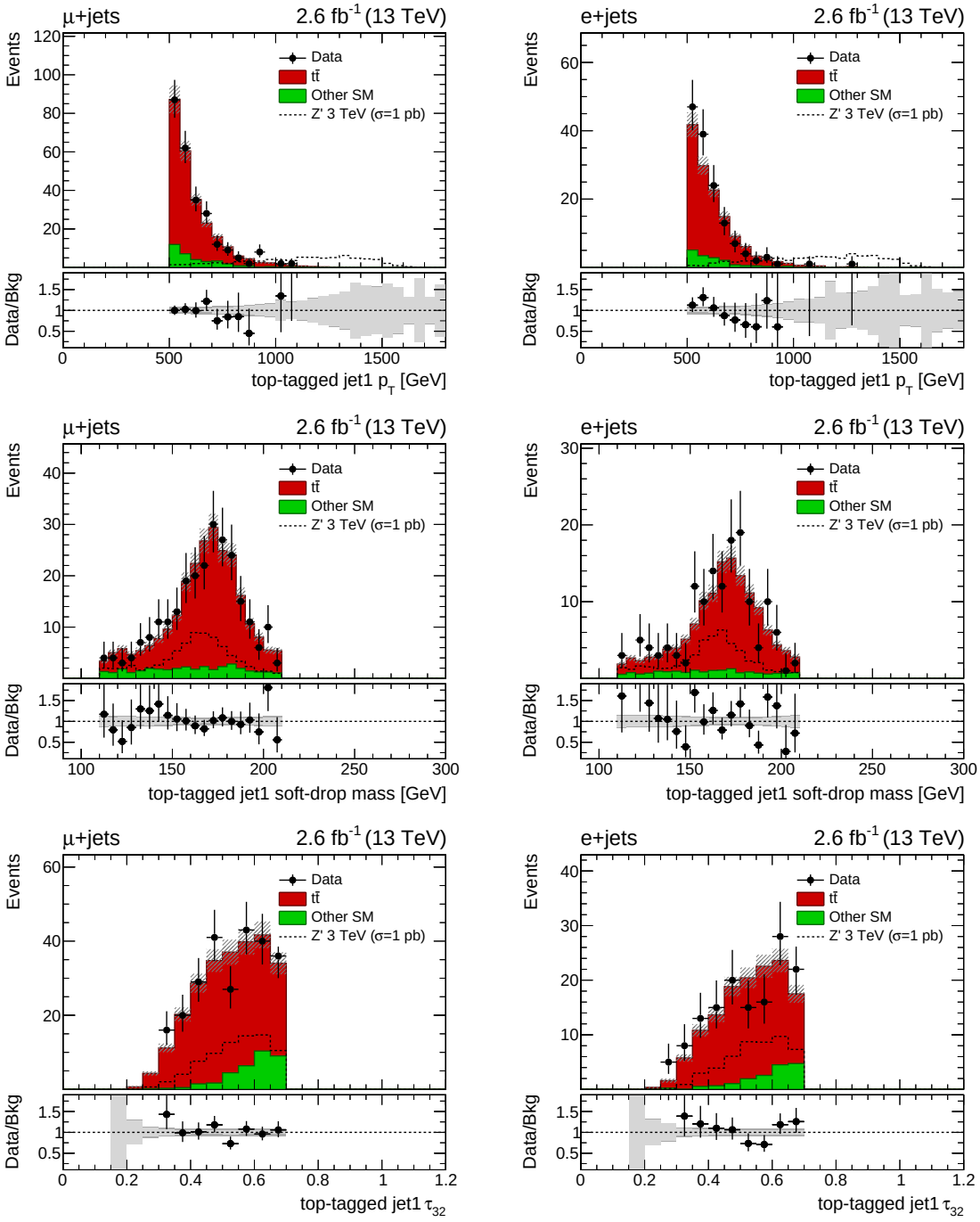


Figure 5.44: Data/MC comparison for the p_T, softdrop mass and τ_{32} ratio of the AK8 jet candidate passing the t-tagging selection, in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

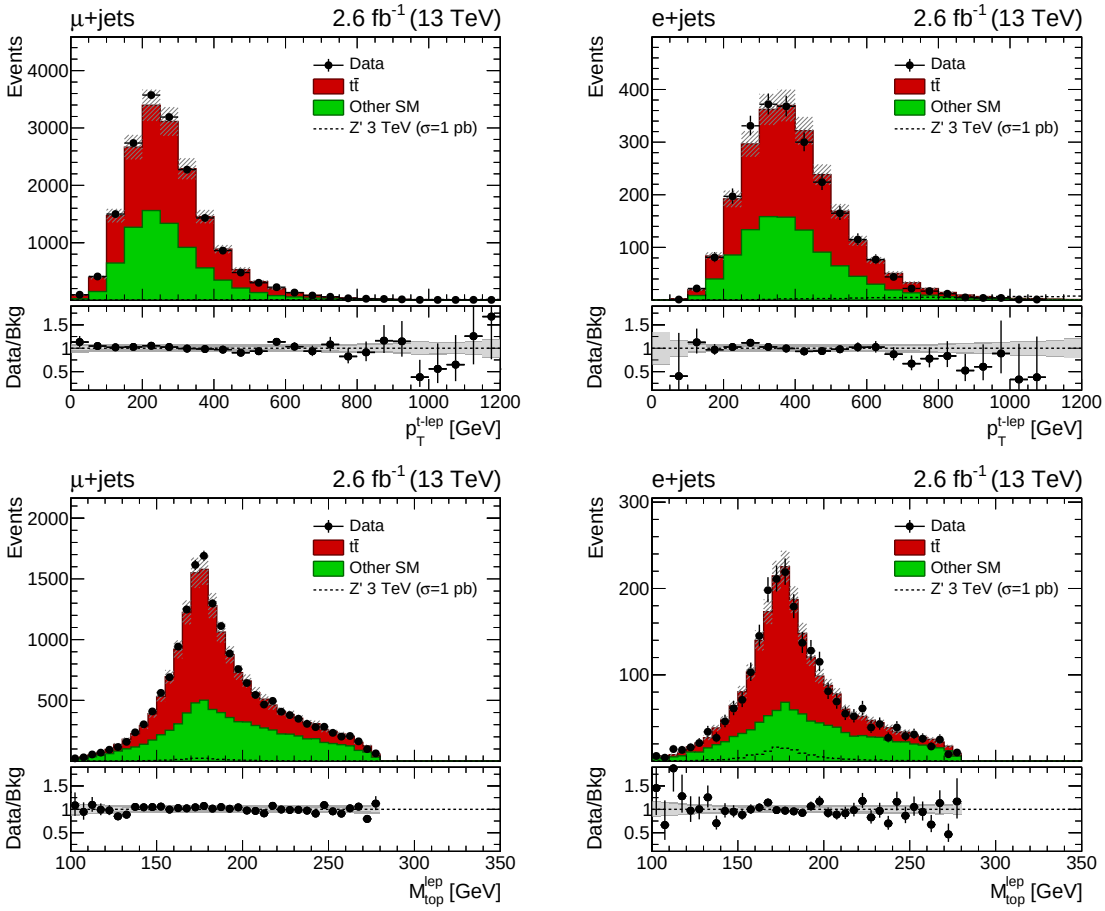


Figure 5.45: Data/MC comparison for the mass and p_T of the reconstructed leptonic top quark in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

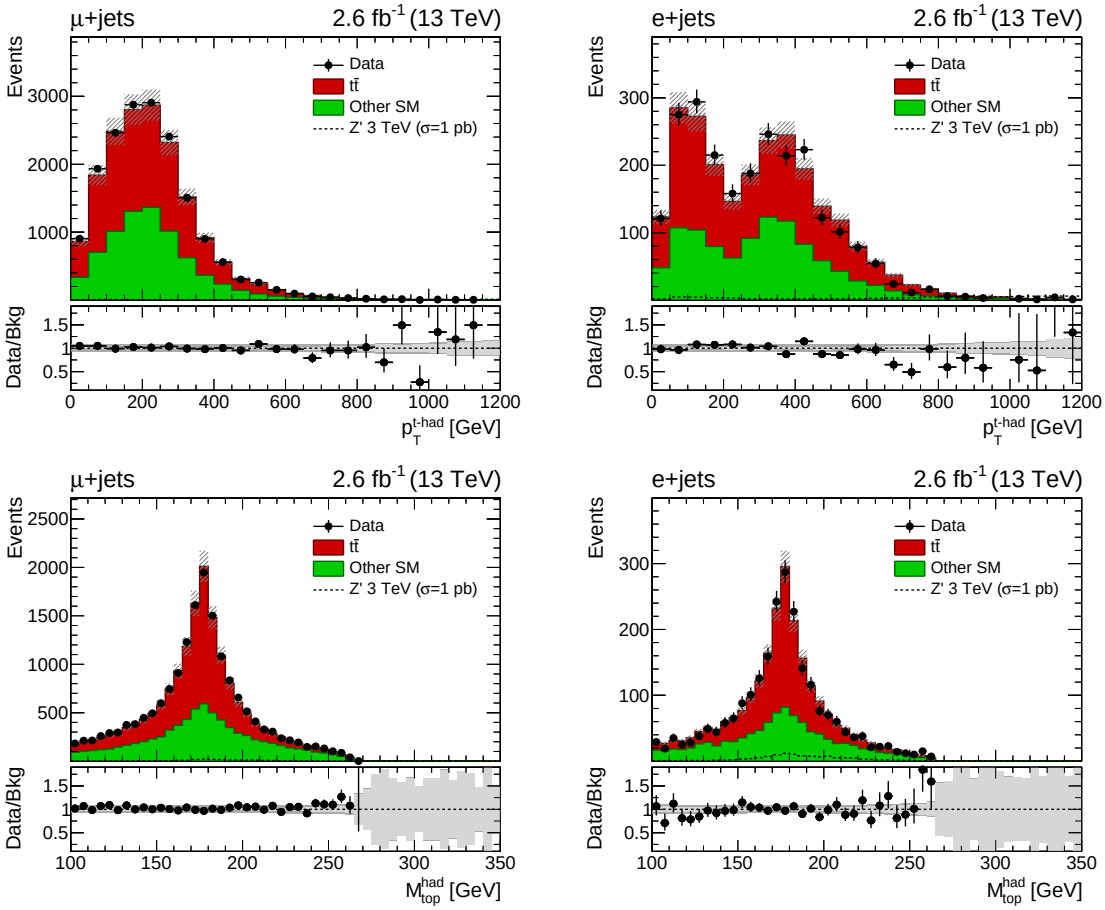


Figure 5.46: Data/MC comparison for the mass and p_T of the reconstructed hadronic top quark in events passing the ℓ + jets SR selection ($\chi^2 < 30$). The background distributions are determined from the likelihood fit described in the text. Plots for the μ + jets (e + jets) channel are shown on the left (right). Each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60. The error associated to the background expectation includes the MC statistical error and the post-fit systematic uncertainties. In the ratio plots, the statistical (light gray) and total uncertainty (dark gray) are shown separately.

5.7 Statistical analysis and results

Since no significant deviation from the SM expectation is observed in the data, we proceed to set exclusion limits on the cross section of a $t\bar{t}$ resonance in different BSM scenarios.

A Bayesian statistical method is used to extract 95% confidence level (CL) upper limits on the cross section times branching ratio, i.e. $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$, for a new massive particle decaying to a $t\bar{t}$ pair. The limit-setting procedure is performed with a template-based statistical evaluation using the invariant mass distribution of the reconstructed $t\bar{t}$ pair. The statistical model is constructed similarly to the one used in the Run-1 version of this analysis, as described in Section 4.8. The THETA package [131] is the software used to perform the statistical analysis.

The invariant mass distributions of the reconstructed $t\bar{t}$ pair in the six ℓ + jets SR categories are used as observables in the statistical analysis. The background distributions are obtained from the maximum-likelihood fit described in Section 5.6. The systematic uncertainties listed in Section 5.5 are introduced as nuisance parameters in the limit calculation. As discussed in Section 5.5, the uncertainty on the t-tagging efficiency correction is left unconstrained (flat prior). The nuisance parameters that affect the shape of the $M_{t\bar{t}}$ distributions are assigned a gaussian prior with mean equal to zero and width equal to unity; the remaining nuisance parameters, which are associated to systematic uncertainties affecting only the normalization of a given process, are assigned a log-normal prior. The statistical uncertainties of the MC prediction are included in the statistical model using the Barlow-Beeston "light" method [132].

Exclusion limits are set for four different types of $t\bar{t}$ resonances: a Z′ boson with relative decay width ($\Gamma_{Z'}/M_{Z'}$) of 1%, 10% and 30%, and a Kaluza-Klein gluon resonance in the RS model. The 95% CL upper limits on $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$ for each of these four BSM signals are shown in Figures 5.47–5.50; the corresponding numerical values are reported in Tables 5.9–5.12. The observed and expected limits on $\sigma \cdot \mathrm{BR}$ are found to in agreement with each other, within the uncertainties of the expected limits, for all the signal mass hypotheses considered. Figure 5.51 includes a comparison between the expected limits for the μ + jets channel, the e + jets channel and the combination.

For both the μ + jets and e + jets channels, the expected sensitivity of this search is driven by the events in the `T1` category, especially for the highest-mass signal hypotheses ($M_X > 1.5$ TeV); the `T0B1` category contributes significantly to the overall sensitivity for lower-mass resonances, while the `T0B0` category remains the one with the lowest signal-over-background ratio and it is used to constrain the uncertainties on non-$t\bar{t}$ backgrounds, e.g. W+jets production. Better exclusion limits are obtained in the μ+jets analysis, compared to the e+jets analysis, across the whole mass spectrum considered. The difference in sensitivity between the two channels is accentuated in the low-mass region ($0.5 \text{ TeV} < M_X < 1.25$ TeV); this is due to the fact that the tighter selection cuts used in the e + jets channel reduce the signal efficiency for the lowest-mass signal hypotheses. As expected, the best overall sensitivity is obtained using the combination of the two channels.

Among the four $X \to t\bar{t}$ models, the most stringent cross section limits are set on the Z′ narrow-width signals; the cross section limits worsen for wider resonances, since the latter are characterized by a broader peak in the $M_{t\bar{t}}$ distributions and an increasing contribution from off-shell production. The limits for the KK gluon signal lie between those of the 10%- and 30%-width Z′ signals, as expected based on the width of these KK resonances.

Based on the theoretical cross sections predicted for these models, these cross section limits are recast into exclusion limits on the mass of the hypothetical $X \to t\bar{t}$ signal. These

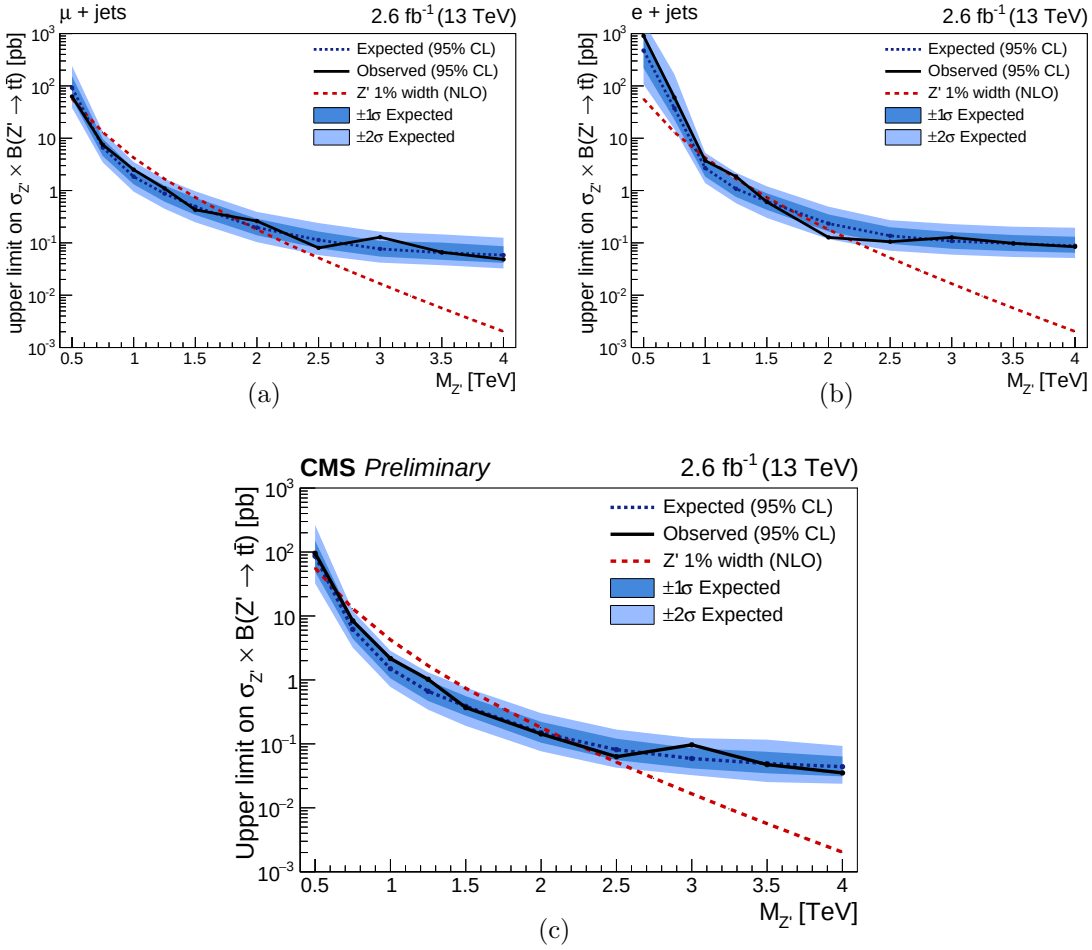


Figure 5.47: 95% CL Bayesian upper limits on the production cross section times branching ratio for a Z′ boson with a relative width ($\Gamma_{Z'}/M_{Z'}$) of 1% decaying to a $t\bar{t}$ pair, shown as a function of the signal mass hypothesis. The theoretical predictions correspond to the QCD-NLO cross sections for a leptophobic Z′ boson in the topcolor model [139]. Limits are shown for the μ+jets (a) and e+jets (b) channels separately and for the statistical combination of the two channels (c) [4].

mass limits are summarized in Table 5.13. The strongest mass limits are the ones obtained for the $X \to t\bar{t}$ signals with the larger width: this is due to the fact that, even though the cross section limits become less stringent as the width of the BSM signal increases, the theoretical cross sections for a $X \to t\bar{t}$ resonance in the BSM scenarios under consideration grow even more rapidly as the signal decay width becomes larger. Taking as a reference the $X \to t\bar{t}$ cross sections calculated at QCD-NLO accuracy for a leptophobic Z′ boson in the topcolor model [139], the experimental data exclude a $Z' \to t\bar{t}$ resonance with 1%-width for masses between 0.64 TeV and 2.25 TeV; a $Z' \to t\bar{t}$ resonance with a relative width of 10% is excluded from 0.5 TeV up to a mass of 3.43 TeV and a $Z' \to t\bar{t}$ resonance with 30%-width is excluded across the whole mass range considered for this model, from 1 TeV to 4 TeV.

Table 5.9: Numerical values for the expected and observed limits shown in Figure 5.47 for the production cross section times branching ratio of a 1%-width Z′ boson decaying to a $t\bar{t}$ pair. Limits are given for each of the two analysis channels and their combination.

$\ell+$jets **combination**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	86.2	150.8 – 49.3	264.8 – 32.4	95.6
0.75	6.2	8.7 – 4.4	12.1 – 3.2	8.4
1.0	1.50	2.07 – 1.06	2.85 – 0.78	2.16
1.25	0.66	0.98 – 0.47	1.32 – 0.34	1.02
1.5	0.39	0.56 – 0.28	0.76 – 0.19	0.37
2.0	0.149	0.221 – 0.104	0.302 – 0.077	0.143
2.5	0.081	0.121 – 0.056	0.167 – 0.042	0.063
3.0	0.059	0.086 – 0.041	0.123 – 0.032	0.097
3.5	0.049	0.075 – 0.035	0.116 – 0.025	0.047
4.0	0.044	0.063 – 0.031	0.093 – 0.024	0.035

$\mu+$jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	92.2	153.1 – 56.1	240.2 – 38.0	62.7
0.75	6.6	9.7 – 4.5	13.1 – 3.4	7.5
1.0	1.84	2.61 – 1.32	3.51 – 0.95	2.51
1.25	0.88	1.25 – 0.61	1.67 – 0.45	1.10
1.5	0.49	0.73 – 0.34	0.97 – 0.25	0.43
2.0	0.196	0.289 – 0.140	0.391 – 0.103	0.261
2.5	0.113	0.165 – 0.079	0.239 – 0.058	0.080
3.0	0.077	0.109 – 0.055	0.163 – 0.042	0.129
3.5	0.065	0.101 – 0.047	0.145 – 0.037	0.066
4.0	0.058	0.086 – 0.041	0.125 – 0.032	0.048

$e+$jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	474.3	902.2 – 220.0	1690.7 – 106.8	903.4
0.75	37.4	72.0 – 22.6	167.5 – 14.9	59.8
1.0	2.7	3.9 – 1.8	5.3 – 1.4	3.7
1.25	1.08	1.60 – 0.75	2.20 – 0.57	1.85
1.5	0.62	0.89 – 0.43	1.20 – 0.30	0.60
2.0	0.23	0.35 – 0.16	0.49 – 0.12	0.13
2.5	0.137	0.199 – 0.096	0.271 – 0.071	0.105
3.0	0.108	0.161 – 0.077	0.229 – 0.060	0.127
3.5	0.097	0.141 – 0.070	0.204 – 0.054	0.098
4.0	0.088	0.131 – 0.065	0.195 – 0.051	0.084

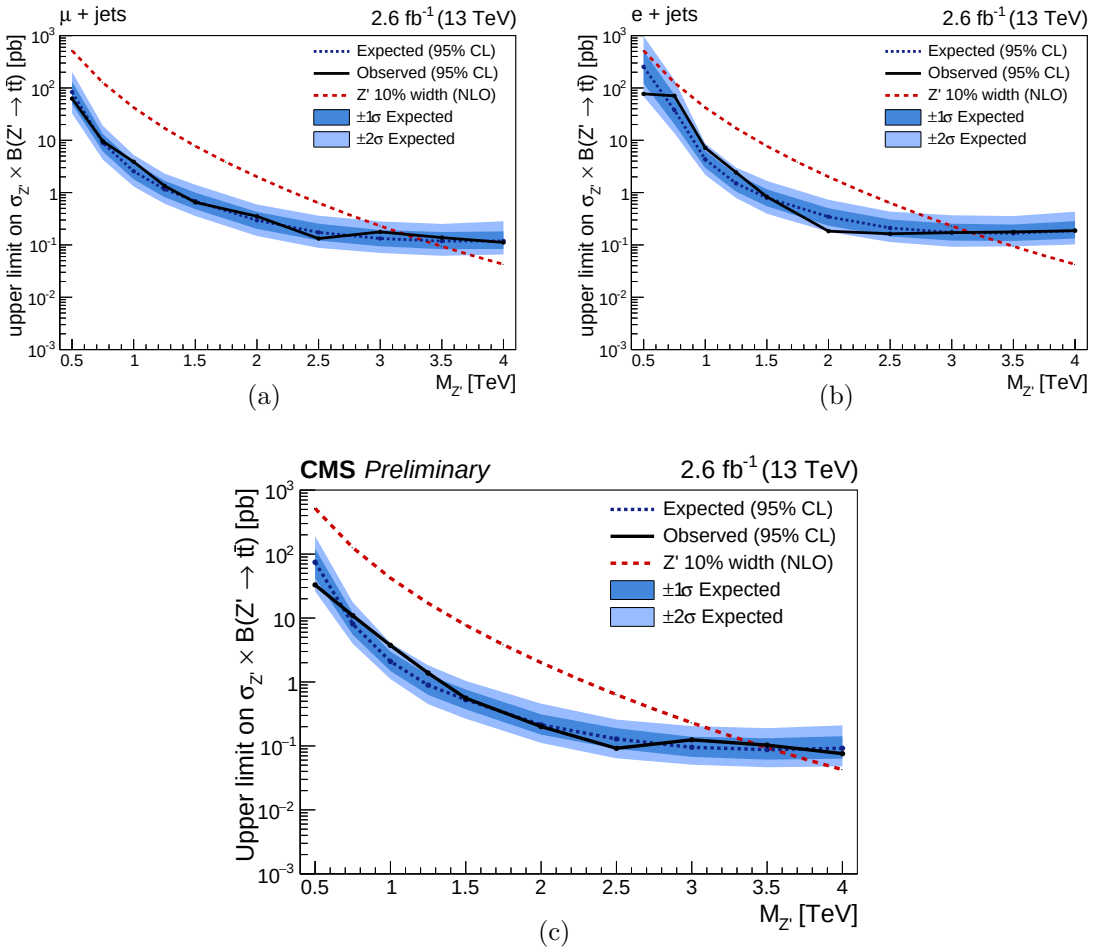


Figure 5.48: 95% CL Bayesian upper limits on the production cross section times branching ratio for a Z′ boson with a relative width ($\Gamma_{Z'}/M_{Z'}$) of 10% decaying to a $t\bar{t}$ pair, shown as a function of the signal mass hypothesis. The theoretical predictions correspond to the QCD-NLO cross sections for a leptophobic Z′ boson in the topcolor model [139]. Limits are shown for the μ + jets (a) and e + jets (b) channels separately and for the statistical combination of the two channels (c) [4].

Table 5.10: Numerical values for the expected and observed limits shown in Figure 5.48 for the production cross section times branching ratio of a 10%-width Z′ boson decaying to a $t\bar{t}$ pair. Limits are given for each of the two analysis channels and their combination.

$\ell+$jets **combination**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	74.6	120.8 – 41.9	191.4 – 26.2	33.0
0.75	8.0	12.0 – 5.5	17.7 – 4.0	10.9
1.0	2.09	2.98 – 1.49	4.07 – 1.11	3.74
1.25	0.89	1.29 – 0.63	1.82 – 0.45	1.38
1.5	0.53	0.76 – 0.37	1.04 – 0.27	0.56
2.0	0.21	0.31 – 0.15	0.46 – 0.11	0.20
2.5	0.129	0.189 – 0.090	0.258 – 0.064	0.092
3.0	0.095	0.139 – 0.067	0.204 – 0.051	0.125
3.5	0.088	0.132 – 0.061	0.190 – 0.046	0.103
4.0	0.092	0.142 – 0.063	0.210 – 0.048	0.076

$\mu+$jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	83.6	132.9 – 49.5	204.4 – 32.5	62.6
0.75	8.9	13.2 – 6.1	18.9 – 4.3	9.6
1.0	2.6	3.8 – 1.8	5.3 – 1.3	3.9
1.25	1.16	1.69 – 0.82	2.31 – 0.61	1.35
1.5	0.69	1.00 – 0.48	1.42 – 0.36	0.65
2.0	0.30	0.43 – 0.20	0.59 – 0.15	0.35
2.5	0.174	0.255 – 0.121	0.362 – 0.088	0.132
3.0	0.133	0.192 – 0.094	0.281 – 0.070	0.177
3.5	0.118	0.176 – 0.083	0.252 – 0.062	0.138
4.0	0.119	0.181 – 0.084	0.284 – 0.066	0.112

$e+$jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	253.2	508.3 – 121.6	974.8 – 68.9	77.1
0.75	37.8	74.5 – 20.3	142.3 – 13.6	70.8
1.0	4.3	6.3 – 3.0	8.7 – 2.2	7.2
1.25	1.50	2.22 – 1.04	3.01 – 0.79	2.43
1.5	0.80	1.16 – 0.55	1.67 – 0.39	0.83
2.0	0.35	0.50 – 0.23	0.73 – 0.18	0.18
2.5	0.211	0.306 – 0.145	0.433 – 0.113	0.163
3.0	0.173	0.254 – 0.121	0.369 – 0.092	0.172
3.5	0.167	0.245 – 0.119	0.356 – 0.094	0.177
4.0	0.188	0.284 – 0.132	0.431 – 0.102	0.187

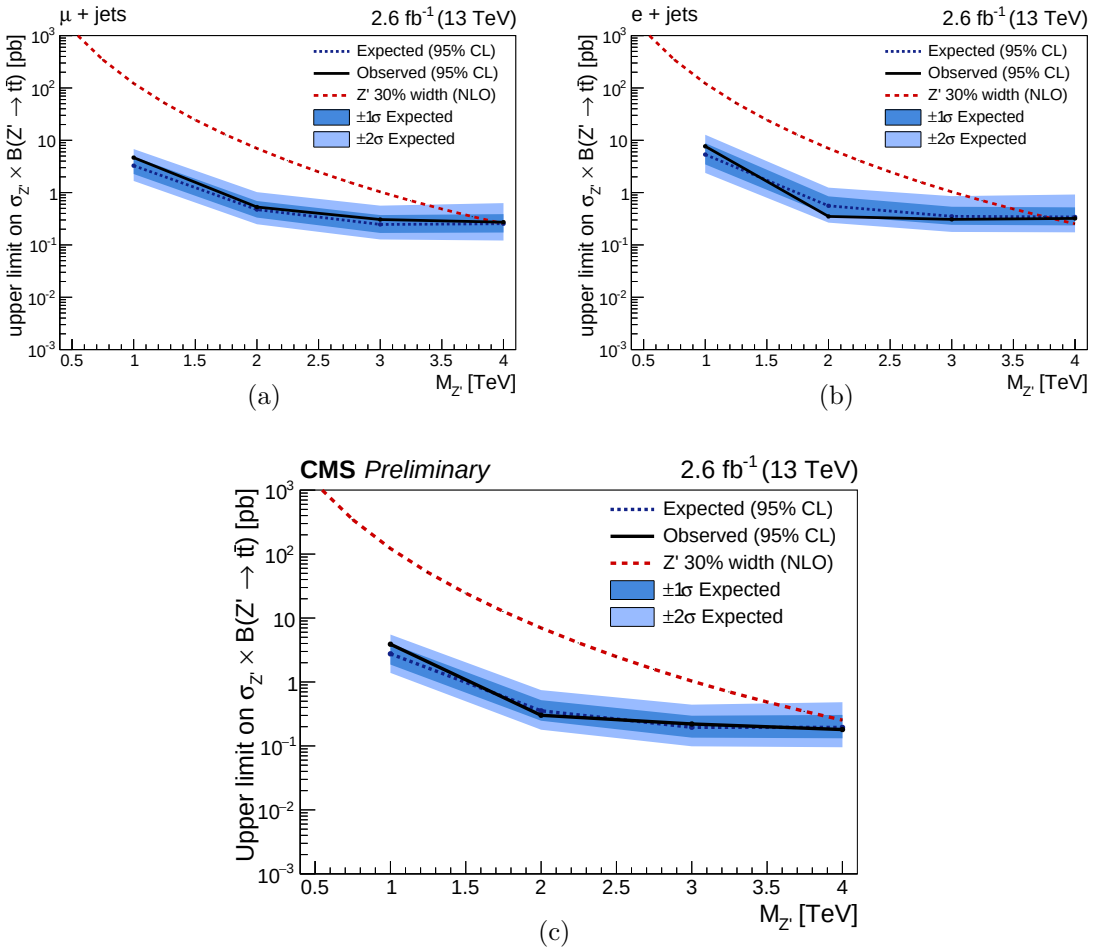


Figure 5.49: 95% CL Bayesian upper limits on the production cross section times branching ratio for a Z′ boson with a relative width ($\Gamma_{Z'}/M_{Z'}$) of 30% decaying to a $t\bar{t}$ pair, shown as a function of the signal mass hypothesis. The theoretical predictions correspond to the QCD-NLO cross sections for a leptophobic Z′ boson in the topcolor model [139]. Limits are shown for the μ + jets (a) and e + jets (b) channels separately and for the statistical combination of the two channels (c) [4].

Table 5.11: Numerical values for the expected and observed limits shown in Figure 5.49 for the production cross section times branching ratio of a 30%-width Z′ boson decaying to a $t\bar{t}$ pair. Limits are given for each of the two analysis channels and their combination.

ℓ + jets **combination**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
1.0	2.8	4.0 – 1.9	5.5 – 1.4	3.9
2.0	0.35	0.52 – 0.25	0.75 – 0.18	0.30
3.0	0.196	0.297 – 0.135	0.442 – 0.099	0.220
4.0	0.195	0.305 – 0.132	0.483 – 0.095	0.180

μ + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
1.0	3.3	4.8 – 2.3	6.8 – 1.7	4.7
2.0	0.47	0.69 – 0.33	1.02 – 0.25	0.53
3.0	0.25	0.37 – 0.17	0.56 – 0.13	0.31
4.0	0.25	0.39 – 0.17	0.63 – 0.12	0.27

e + jets **channel**

$M_{Z'}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
1.0	5.3	8.8 – 3.5	12.8 – 2.4	7.7
2.0	0.56	0.85 – 0.37	1.24 – 0.27	0.35
3.0	0.35	0.54 – 0.24	0.86 – 0.18	0.31
4.0	0.34	0.52 – 0.24	0.92 – 0.17	0.32

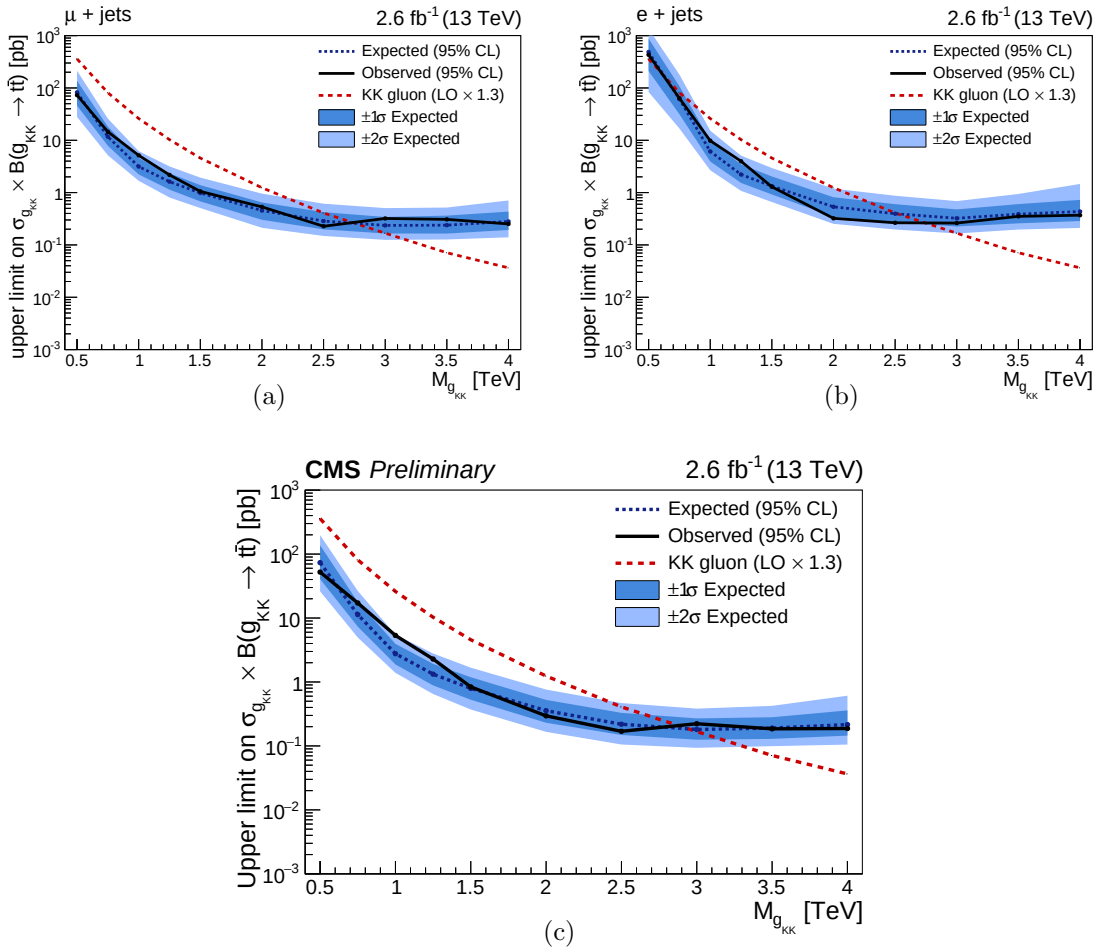


Figure 5.50: 95% CL Bayesian upper limits on the production cross section times branching ratio for a KK gluon decaying to a $t\bar{t}$ pair, shown as a function of the signal mass hypothesis. The theoretical predictions for the cross section of a KK gluon in the RS model are taken from PYTHIA-v8.2 and multiplied by a factor $K = 1.3$ in order to account for higher-order corrections [113]. Limits are shown for the μ + jets (a) and e + jets (b) channels separately and for the statistical combination of the two channels (c) [4].

Table 5.12: Numerical values for the expected and observed limits shown in Figure 5.50 for the production cross section times branching ratio of a KK gluon resonance decaying to a $t\bar{t}$ pair. Limits are given for each of the two analysis channels and their combination.

$\ell+$jets **combination**

$M_{g_{\mathrm{KK}}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	73.4	137.1 – 40.7	197.4 – 26.1	52.3
0.75	11.3	17.9 – 7.3	26.5 – 5.0	17.2
1.0	2.8	4.0 – 1.9	5.5 – 1.4	5.3
1.25	1.33	1.92 – 0.88	2.79 – 0.65	2.28
1.5	0.79	1.19 – 0.53	1.67 – 0.37	0.84
2.0	0.36	0.52 – 0.23	0.76 – 0.17	0.30
2.5	0.22	0.33 – 0.15	0.47 – 0.11	0.17
3.0	0.181	0.269 – 0.125	0.383 – 0.093	0.222
3.5	0.190	0.281 – 0.129	0.423 – 0.099	0.185
4.0	0.22	0.36 – 0.15	0.61 – 0.11	0.19

$\mu+$jets **channel**

$M_{g_{\mathrm{KK}}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	82.4	136.8 – 45.9	212.7 – 28.0	72.8
0.75	11.7	18.2 – 7.8	26.0 – 5.2	14.6
1.0	3.1	4.6 – 2.2	6.2 – 1.7	5.2
1.25	1.62	2.34 – 1.14	3.17 – 0.81	2.18
1.5	0.98	1.39 – 0.68	1.93 – 0.51	1.05
2.0	0.45	0.65 – 0.30	0.96 – 0.21	0.53
2.5	0.29	0.42 – 0.20	0.61 – 0.15	0.23
3.0	0.24	0.34 – 0.17	0.51 – 0.12	0.32
3.5	0.24	0.36 – 0.16	0.52 – 0.13	0.31
4.0	0.28	0.43 – 0.19	0.71 – 0.14	0.25

$e+$jets **channel**

$M_{g_{\mathrm{KK}}}$ [TeV]	Expected [pb]	Exp. $\pm1\sigma$ [pb]	Exp. $\pm2\sigma$ [pb]	Observed [pb]
0.5	487.0	880.0 – 210.4	1351.4 – 83.7	430.6
0.75	61.3	107.9 – 31.5	180.3 – 16.8	63.7
1.0	6.1	9.7 – 3.9	15.3 – 2.7	9.8
1.25	2.2	3.5 – 1.5	5.1 – 1.1	4.0
1.5	1.33	2.04 – 0.92	2.95 – 0.66	1.29
2.0	0.54	0.82 – 0.34	1.19 – 0.25	0.32
2.5	0.40	0.60 – 0.27	0.87 – 0.20	0.26
3.0	0.32	0.48 – 0.23	0.69 – 0.17	0.26
3.5	0.39	0.60 – 0.26	0.94 – 0.20	0.35
4.0	0.43	0.73 – 0.29	1.46 – 0.21	0.37

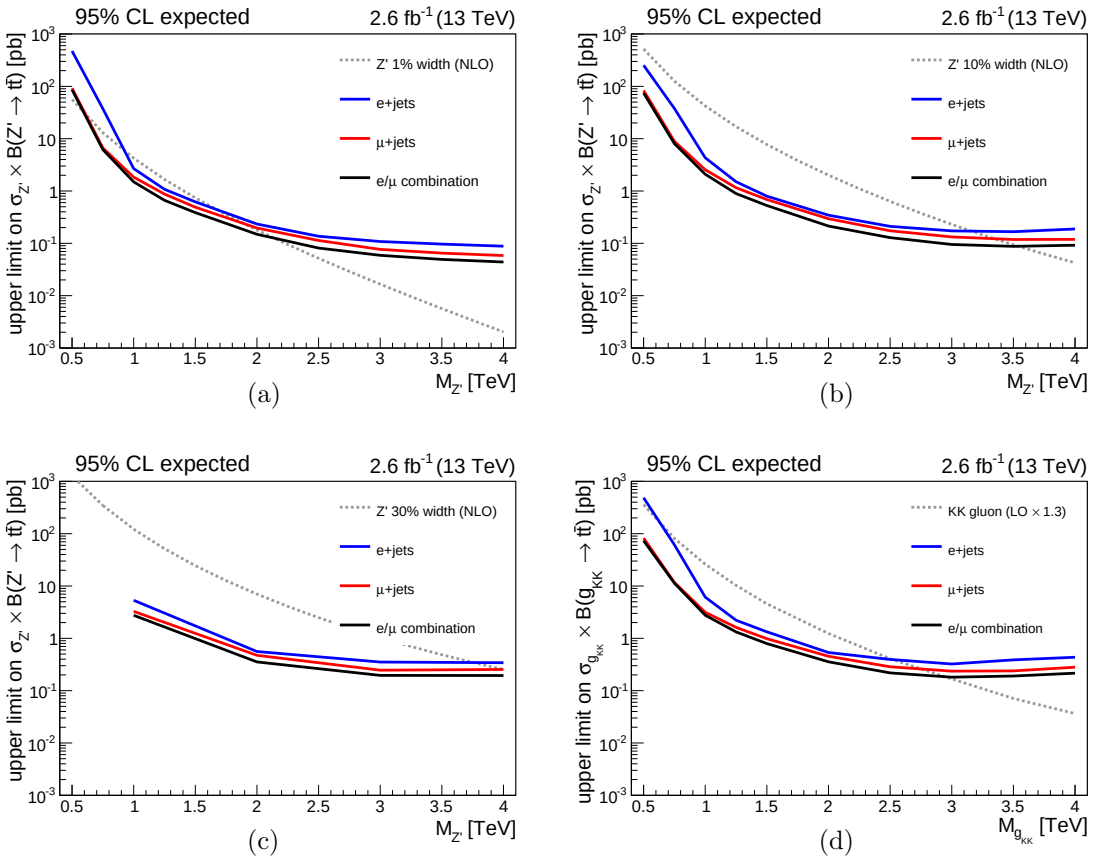


Figure 5.51: Comparison of the 95% CL expected upper limits on $\sigma(pp \to X) \times \mathrm{BR}(X \to t\bar{t})$ for the $\mu+$jets channel, the $e+$jets channel and their combination. Limits are shown for four BSM scenarios: a Z′ boson with 1% width (a), 10% width (b) and 30% width (c) and KK gluon resonance in the RS model (d).

Table 5.13: Observed and expected excluded mass ranges for the BSM $X \to t\bar{t}$ signals considered in the statistical analysis. Mass limits are shown for the $\mu+$jets and $e+$jets channels separately and for their combination.

Excluded mass range [TeV]

signal model	$\mu+$jets channel observed (expected)	$e+$jets channel observed (expected)	e/μ combination observed (expected)
Z′ (1% width)	0.54 – 1.79 (0.61 – 1.90)	0.98 – 1.14, 1.34 – 2.16 (0.92 – 1.68)	0.64 – 2.25 (0.59 – 2.14)
Z′ (10% width)	0.50 – 3.19 (0.50 – 3.35)	0.50 – 3.15 (0.50 – 3.16)	0.50 – 3.43 (0.50 – 3.54)
Z′ (30% width)	1.00 – 3.94 (1.00 – 4.00)	1.00 – 3.82 (1.00 – 3.76)	1.00 – 4.00 (1.00 – 4.00)
KK gluon	0.50 – 2.74 (0.50 – 2.75)	0.61 – 2.75 (0.63 – 2.52)	0.50 – 2.88 (0.50 – 2.95)

Chapter 6

Studies on angular observables and interference effects in $X \to t\bar{t}$ searches

This chapter presents a set of original studies focused on improvements in the search for a $X \to t\bar{t}$ resonance at the LHC. These studies are concerned with the use of the angular observables of the $t\bar{t}$ system and the treatment of interference effects between a $X \to t\bar{t}$ BSM signal and the production of a $t\bar{t}$ pair in the SM. Both of these studies are carried out in the context of the $X \to t\bar{t} \to \ell + \text{jets}$ search at $\sqrt{s} = 13$ TeV described in Chapter 5. The improvements considered in these studies are also tested using projections of the sensitivity of the $\ell + \text{jets}$ analysis at the higher integrated luminosities expected for the next phases of the LHC program.

6.1 Angular observables of the $t\bar{t}$ system

This section presents a series of studies based on the use of the angular observables of the $t\bar{t}$ system as a tool to complement the measurement of the $M_{t\bar{t}}$ spectra in the search for a $t\bar{t}$ resonance. These studies show how the measurement of angular variables related to the top quark helicity and the production mechanism of the $t\bar{t}$ pair can be used to improve the sensitivity of the analysis and characterize the properties of a hypothetical $X \to t\bar{t}$ signal.

The section is structured as follows. We first focus on the kinematics of the top quark decay and investigate how the helicity of a top quark originating from a $X \to t\bar{t}$ signal can affect the results of the $\ell + \text{jets}$ analysis; we show that different top quark polarizations lead to important differences in the kinematic distributions of the top quark decay products and, as a consequence, the efficiency of the $\ell + \text{jets}$ analysis selection differs for top quarks with left-handed (LH) and right-handed (RH) helicities. Secondly, we consider another angular distribution, the so-called top quark $\cos\theta^*$ distribution: this quantity is sensitive to the production mechanism of the $t\bar{t}$ pair. The final part of this section contains the exclusion limits for $X \to t\bar{t}$ signals characterized by different top quark polarizations and the projections of the expected sensitivity of the $\ell + \text{jets}$ analysis for higher integrated luminosities; in this context, we consider one example of how to use top quark helicity observables to improve the sensitivity of the analysis.

6.1.1 Comparison of $Z' \to t\bar{t}$ signals with different top quark helicities

We first study how the different helicities of the top quarks from a $X \to t\bar{t}$ signal can have an impact on the results on the ℓ + jets analysis.

To this end, we generate $X \to t\bar{t}$ MC samples enriched with top quarks with left-handed helicity, in one case, and right-handed helicity, in the other. This is done by exploiting the fact that, in the case of $X \to t\bar{t}$ signals with very high masses ($M_X \gtrsim 1$ TeV), the top quarks in the final state are produced with an average momentum that largely exceeds the top quark mass and their helicity is thus strongly correlated to the chiral couplings (g_L, g_R) of the new state X to the top quark. Based on this, we produce MC samples for the production of a $Z' \to t\bar{t}$ signal with 1% width at $\sqrt{s} = 13$ TeV using `MadGraph5_aMC@NLO` at LO for two opposite choices of the effective couplings of the Z′ boson to the top quark: (1) the purely left-handed case, given by $g_L = 1$ and $g_R = 0$, and (2) the purely right-handed case, given by $g_L = 0$ and $g_R = 1$. In addition to these, we consider a third case given by the $Z' \to t\bar{t}$ signal samples used in Chapter 4 and Chapter 5. In these MC samples, the Z′ boson is assigned the same couplings to fermions as the SM Z boson (SM-like case); in the case of the coupling to the top quark, this implies that the absolute value of the g_L coupling is much bigger than that of the g_R coupling. Since the direct relation between the values of the Z′ couplings and the top quark helicities holds only in the high-energy limit ($E_t \gg m_t$), only $Z' \to t\bar{t}$ signals with $M_{Z'} \geqslant 1$ TeV are considered in this study; notably, these high-mass signals are also the most relevant ones for future $X \to t\bar{t}$ searches. Based on the above, we effectively work in the limit in which the signal samples with purely LH (RH) couplings contain only top quarks with left-handed (right-handed) helicity; in what follows, the terms left-handed and right-handed refer only to the top quark helicities. It is important to note that, although this study considers the $Z' \to t\bar{t}$ signal, the effects due to the top quark polarization do not depend on the specific type of spin-1 $X \to t\bar{t}$ resonance considered.

Figure 6.1 shows a comparison between the three different types of $Z' \to t\bar{t}$ MC samples defined above, considering some relevant kinematic quantities at generator level. When considering the MC simulation with RH top quarks, we find that, compared to the LH case, the average value of the lepton p_T is expected to be higher, while the p_T spectra of the neutrino and b-quark are expected to be softer. As expected, the distributions obtained in the SM-like and LH cases are found to be very similar to each other.

Figure 6.2 shows a comparison of the signal efficiencies for the ℓ + jets analysis at reconstruction level. As shown in Figure 6.2(a), the $Z' \to t\bar{t}$ signal with RH top quarks shows a higher efficiency overall: this gain is driven by the larger signal acceptance due to the harder lepton p_T spectrum. The cuts on $\not{E}_T$ and jet p_T (which favor the LH case) reduce this difference in efficiency only partially, because the original differences in the neutrino and b-quark energy spectra are smeared at reconstruction level by the $\not{E}_T$ and jet energy resolutions. This holds true for all the signal mass hypotheses considered, between 1 TeV and 4 TeV, as can be seen in Figure 6.2(b); a slight reduction for the difference in efficiency between the LH and RH cases can be seen for the highest mass hypotheses, as the average p_T of the lepton increases and the impact of the lepton-p_T cut becomes less important. If we consider the total signal efficiency of the ℓ + jets selection on all generated $Z' \to t\bar{t}$ events (see Figure 6.2), the relative difference in efficiency between the RH case and the SM-like case, calculated as $(\varepsilon_{\mathrm{RH}} - \varepsilon_{\mathrm{SM}})/\varepsilon_{\mathrm{SM}}$, varies from $38\% \pm 10\%$ for $M_{Z'} = 1.5$ TeV to $18\% \pm 7\%$ for $M_{Z'} = 3.5$ TeV. As expected, the same difference in efficiency is much smaller in the LH case: in general, the efficiencies for the LH and SM-like samples are found to be compatible within their statistical uncertainties.

The differences in signal efficiency induced by the top quark helicity, which have been de-

termined using the $Z' \to t\bar{t}$ signal, are expected to hold for a generic spin-1 $X \to t\bar{t}$ resonance. In particular, this effect is relevant to the case of a KK gluon resonance in the RS model. In the standard parameterization of the model, the KK gluon signal is characterized by an enhanced production of right-handed top quarks, as detailed in Section 6.2. On the other hand, the KK gluon MC sample used in the ℓ + jets analysis is generated with PYTHIA-v8.2, which does not take into account the polarization of the top quarks produced in the final state. As a consequence, the top quarks in these signal samples are treated as unpolarized, which is equivalent to having the same proportion of left-handed and right-handed top quarks in the final state; in addition, the spin correlations between the top and antitop quark decays are not taken into account. The above results suggest that, if the top quark polarization were correctly accounted for in the $g_{KK} \to t\bar{t}$ simulation, the efficiency for this signal would be higher, thanks to the larger fraction of right-handed top quarks.

6.1.2 Top quark helicity observables

In the decay of a polarized top quark, the following relation holds for the the angular distribution of its decay products [148]

$$\frac{1}{\Gamma_t}\frac{d\Gamma_t}{d\cos\theta_x} = \frac{1}{2}\left(1 + \mathcal{P}\cdot\alpha_x\cdot\cos\theta_x\right) \tag{6.1}$$

where Γ_t is the total top quark decay width and x denotes any of the top quark decay products ($x = b, f_u, f_d$), where b is the b-quark from the top quark decay and f_u (f_d) is the up-type (down-type) fermion from the W boson decay; $\mathcal{P}$ corresponds to the value of the top quark helicity ($\mathcal{P} = +1$ for right-handed helicity, $\mathcal{P} = -1$ for left-handed helicity), θ_x is the angle between the momentum of the x decay product in the top quark rest frame and the top quark polarization axis; in the helicity basis, the latter coincides with the momentum of the top quark in the $t\bar{t}$ rest frame. The parameter α_x (spin analyzing power) quantifies the correlation between the top quark polarization and the angular distribution of the x decay product; at LO, the numerical values of these parameters are $\alpha_b = -0.4$, $\alpha_{f_u} = -0.3$ and $\alpha_{f_d} = +1$ [148].

We consider the $\cos\theta_x$ distributions for the products of the semileptonic top quark decay, $t \to bW \to b\ell\nu$ with $f_u = \nu$ and $f_d = \ell$; this is the top quark decay mode for which the three different decay products can be identified more easily at the reconstruction level. Figures 6.3–6.5 show the $\cos\theta$ distributions for the b-jet, lepton and neutrino associated to the reconstructed leptonic top quark in the three event categories of the ℓ + jets SR. These plots include a shape comparison between the different $Z' \to t\bar{t}$ MC samples and a comparison between the data and the expected SM background. As expected from the spin analyzing power of the products of the semileptonic top quark decay, the angular variable with the best discrimination power between different top quark polarizations is the lepton $\cos\theta$ distribution. When comparing to the data, we find that the MC simulations used in the analysis reproduce the shape of these angular variables reasonably well in each of the categories of the ℓ + jets SR; data and background distributions are found to be in agreement within the statistical uncertainty of the data. It is worth noting that, since the average polarization for the top quarks from SM $t\bar{t}$ production is very small, these angular distributions are expected to be flat before any event selection is applied; therefore, the data/MC distributions in Figures 6.3–6.5 also give a measure of how these angular variables are sculpted by the analysis cuts.

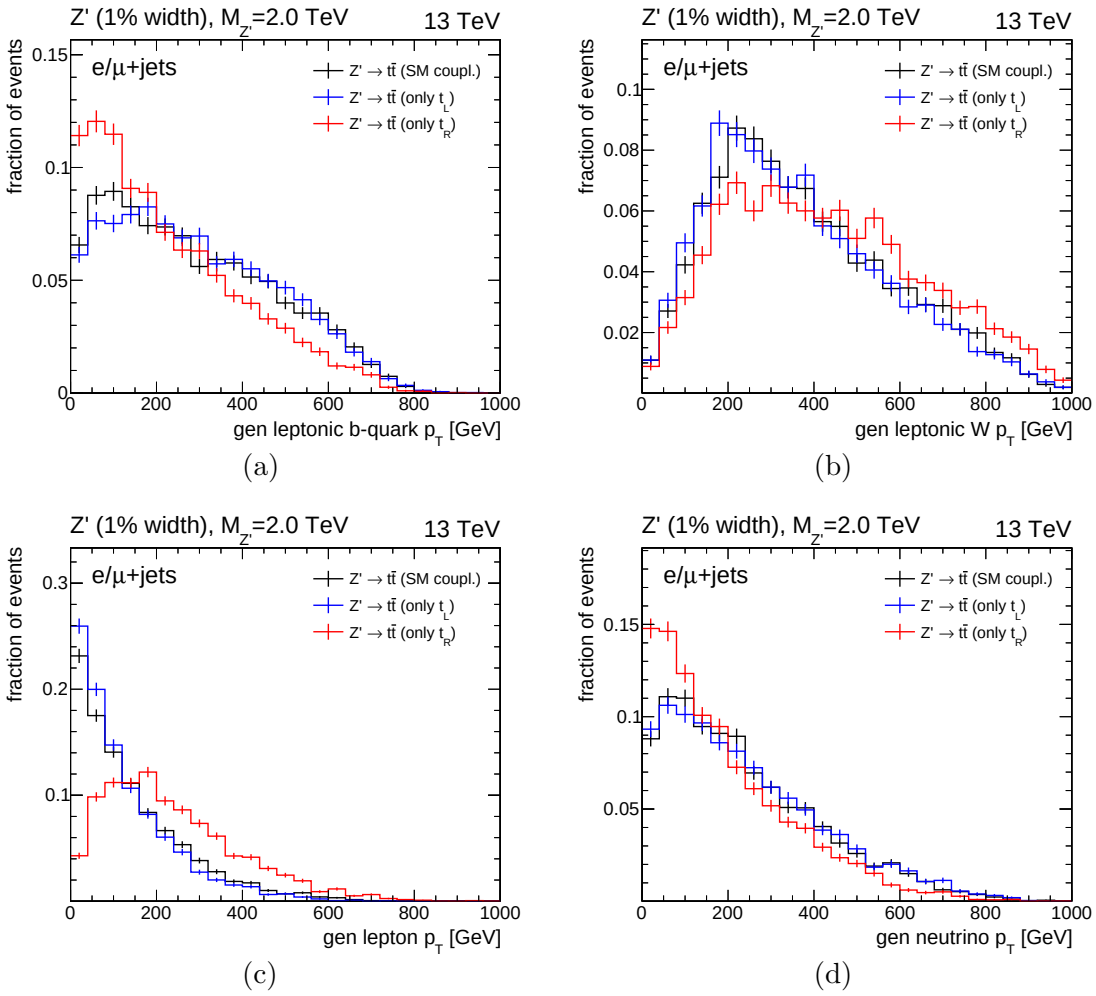


Figure 6.1: Parton-level distributions, before any event selection, for the decay products of the leptonic top quark decay in $Z' \to t\bar{t} \to \ell + \text{jets}$ events with different Z′ couplings to the top quark: (a) b-quark p_T, (b) leptonic-W p_T, (c) lepton p_T, (d) neutrino p_T. The plots are shown for a Z′ boson with $M_{Z'} = 2$ TeV. Each distribution is normalized to unity.

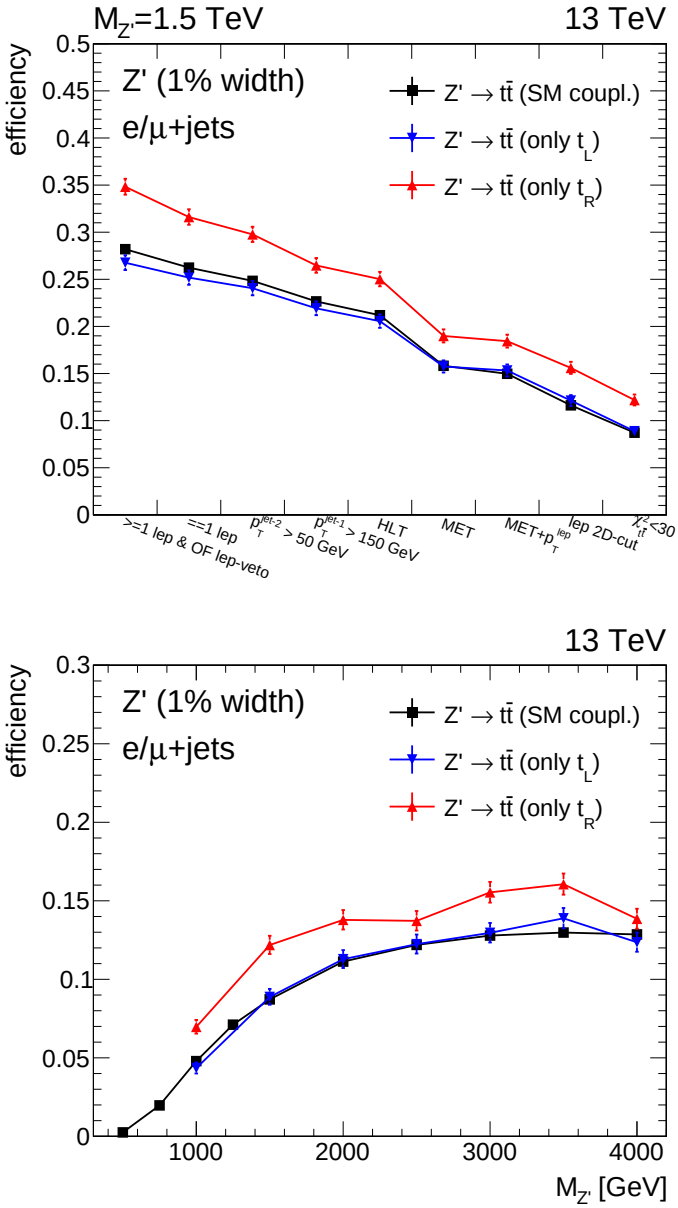


Figure 6.2: Top: overall signal efficiency of the ℓ + jets analysis (at reconstruction level) as a function of the analysis cuts for $Z' \to t\bar{t}$ signals with $M_{Z'} = 1.5$ TeV; the cuts are ordered from left to right and applied sequentially (each on top of the preceding ones). Bottom: overall signal efficiency after the $\chi^2 < 30$ requirement, as a function of the Z' signal mass. In both plots, the efficiencies are measured with respect to all generated events, which include all possible $t\bar{t}$ decay modes (not only ℓ + jets), and the efficiencies for the μ + jets and e + jets channels are added together.

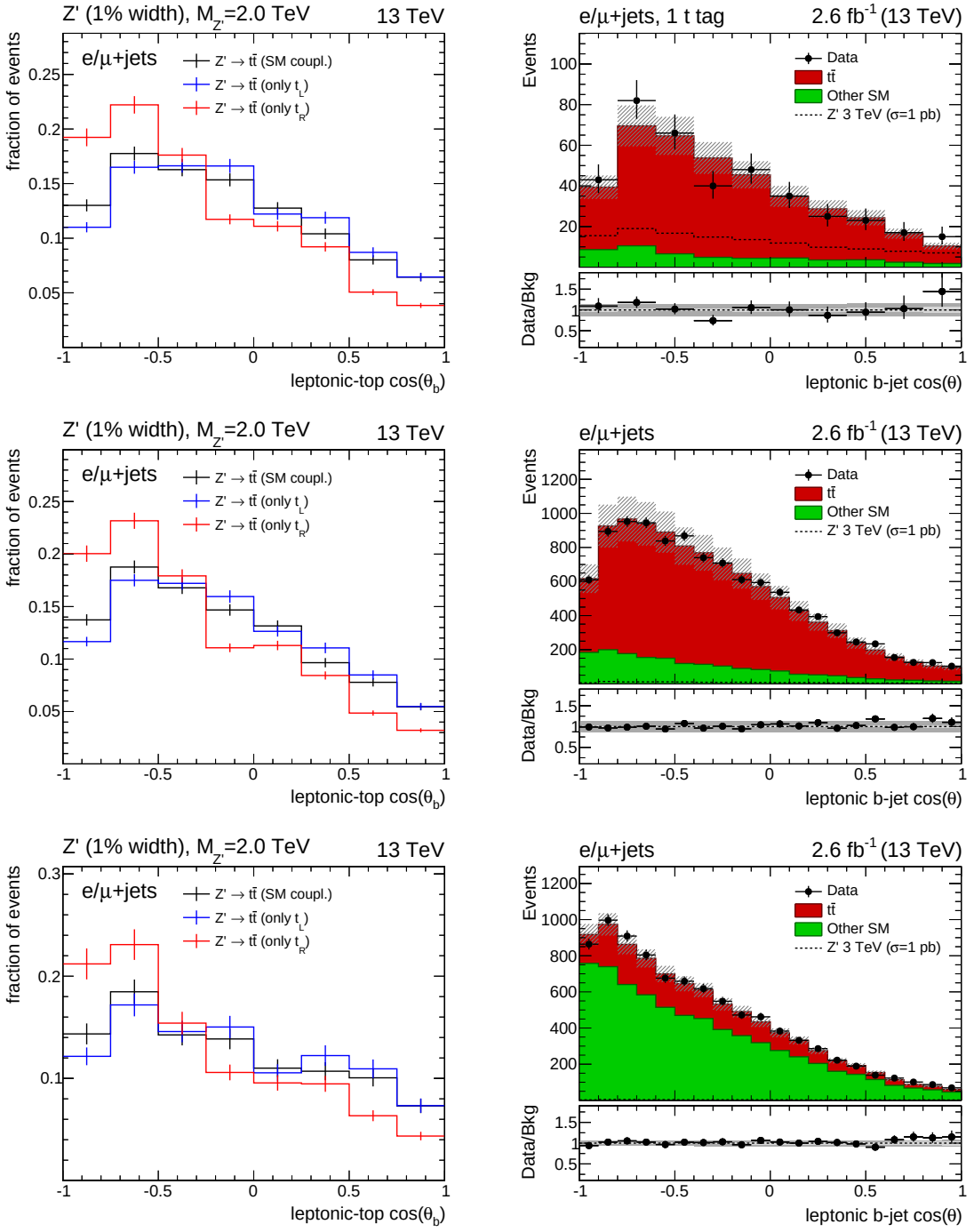


Figure 6.3: Reconstructed $\cos\theta$ distribution for the AK4 jet associated to the leptonic top quark in ℓ + jets events. The three event categories of the ℓ + jets SR are shown separately: (top) `T1` category, (center) `T0B1` category, (bottom) `T0B0` category. Each distribution includes the events of both the μ + jets and e + jets channels. Left: comparison for $Z' \to t\bar{t}$ samples with different top quark polarizations ($M_{Z'} = 2$ TeV), with each distribution normalized to unity. Right: data/MC comparison for events in the ℓ + jets SR; the SM backgrounds are obtained from the likelihood fit described in Section 5.6 and each plot includes the distribution expected for a narrow-width Z′ signal with $M_{Z'} = 3$ TeV (with SM-like couplings) normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60.

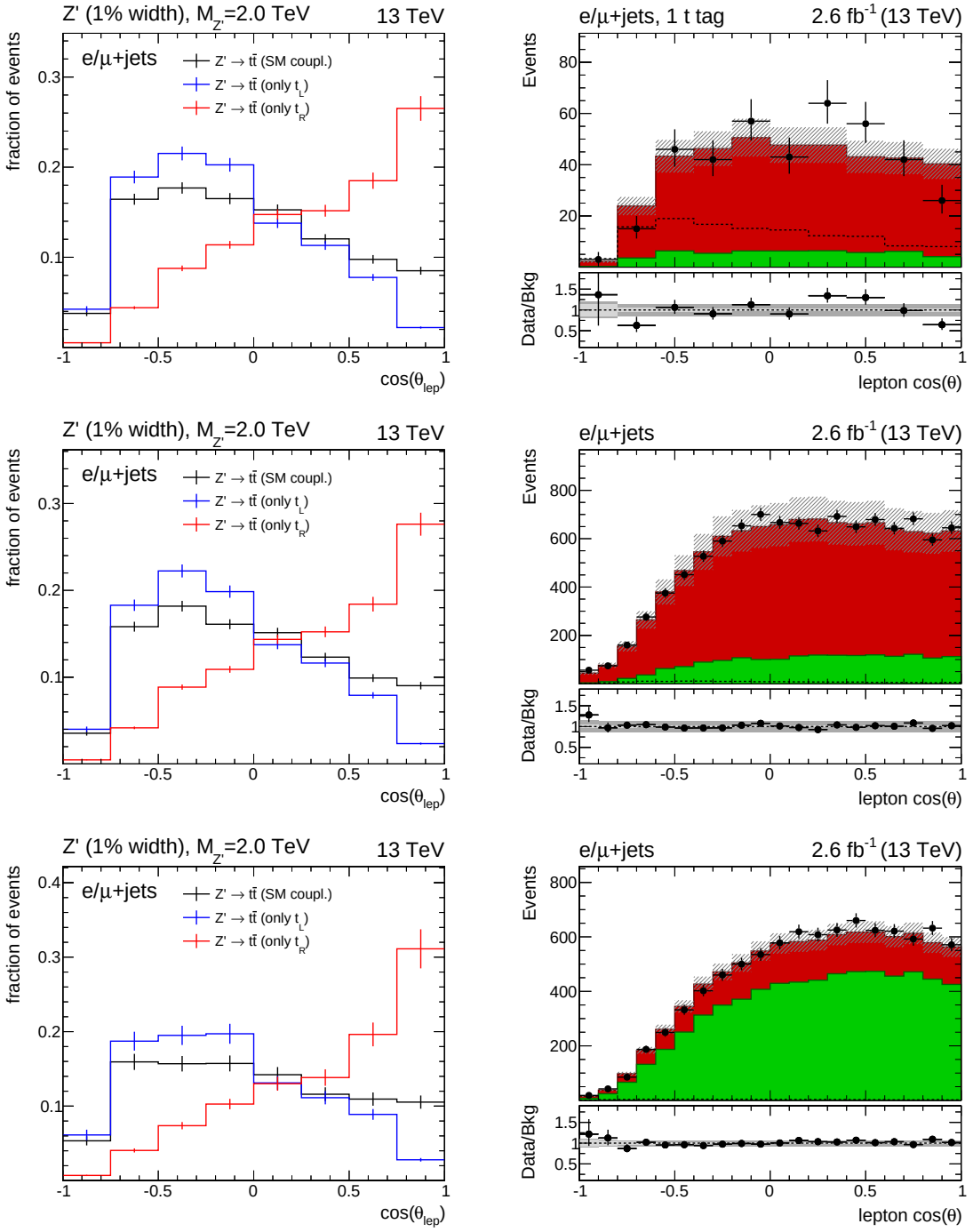


Figure 6.4: Reconstructed $\cos\theta$ distribution for the charged lepton in $\ell+$jets events. The three event categories of the $\ell+$jets SR are shown separately: (top) `T1` category, (center) `T0B1` category, (bottom) `T0B0` category. Each distribution includes the events of both the $\mu+$jets and $e+$jets channels. Left: comparison for $Z' \to t\bar{t}$ samples with different top quark polarizations ($M_{Z'} = 2$ TeV), with each distribution normalized to unity. Right: data/MC comparison for events in the $\ell+$jets SR; the SM backgrounds are obtained from the likelihood fit described in Section 5.6 and each plot includes the distribution expected for a narrow-width Z' signal with $M_{Z'} = 3$ TeV (with SM-like couplings) normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60; the plot legend is the same as the one used in Figure 6.3.

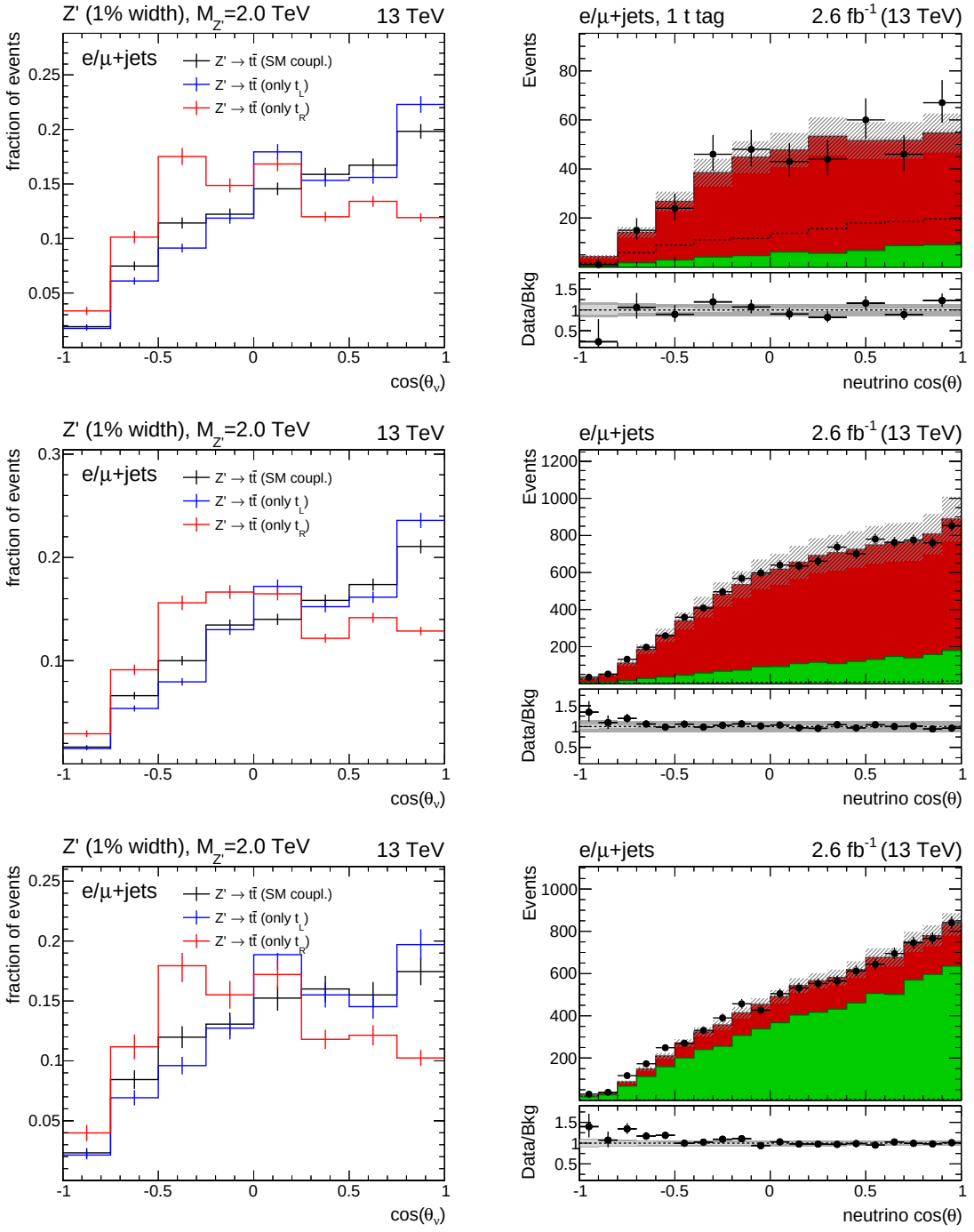


Figure 6.5: Reconstructed $\cos\theta$ distribution for the neutrino component of the $t\bar{t}$ system in ℓ + jets events. The three event categories of the ℓ + jets SR are shown separately: (top) `T1` category, (center) `T0B1` category, (bottom) `T0B0` category. Each distribution includes the events of both the μ + jets and e + jets channels. Left: comparison for $Z' \to t\bar{t}$ samples with different top quark polarizations ($M_{Z'} = 2$ TeV), with each distribution normalized to unity. Right: data/MC comparison for events in the ℓ + jets SR; the SM backgrounds are obtained from the likelihood fit described in Section 5.6 and each plot includes the distribution expected for a narrow-width Z' signal with $M_{Z'} = 3$ TeV (with SM-like couplings) normalized to a cross section of 1 pb, which corresponds to the theoretical cross section times a factor 60; the plot legend is the same as the one used in Figure 6.3.

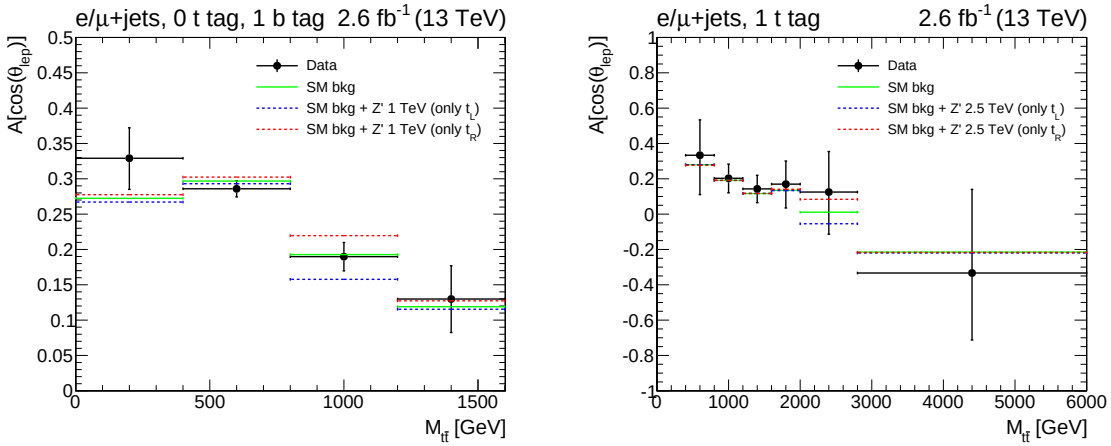


Figure 6.6: Average $\cos\theta_\ell$ polarization asymmetry as a function of the reconstructed $t\bar{t}$ mass: (left) `T0B1` category, (right) `T1` category. In both plots, the asymmetry in data is compared to the expected asymmetry from SM backgrounds and a model with background plus a Z′ signal with LH or RH top quarks; the signal contribution in the left (right) plot corresponds to a $Z' \to t\bar{t}$ resonance with $M_{Z'} = 1$ (2.5) TeV, normalized to the theoretical cross section for a leptophobic Z′ boson in the topcolor model. Each distribution includes the events of both the $\mu + \text{jets}$ and $e + \text{jets}$ channels.

6.1.3 Polarization asymmetry

One method to utilize the $\cos\theta_x$ $(x = b, f_u, f_d)$ distributions in the study of the top quark polarization is to determine the so-called *polarization asymmetry* [24]. We define it as

$$A[\cos\theta_x] = \frac{N[\cos\theta_x > 0] - N[\cos\theta_x < 0]}{N[\cos\theta_x > 0] + N[\cos\theta_x < 0]} \tag{6.2}$$

Since the lepton $\cos\theta_\ell$ distribution is the one that provides the best discrimination between the LH and RH cases, we focus on the asymmetry associated to this angular variable. Figure 6.6 shows the average $\cos\theta_\ell$ polarization asymmetry as a function of the invariant mass of the reconstructed $t\bar{t}$ system for the two most sensitive categories of the $\ell + \text{jets}$ SR, given by events with one b-tagged jet (`T0B1`) and events with one t-tagged jet (`T1`). The SM background prediction obtained from MC is in agreement with the values measured in data, within the statistical uncertainty of the latter. These plots also show the discrimination power provided by the $\cos\theta_\ell$ asymmetry in distinguishing between a $Z' \to t\bar{t}$ signal with LH and RH top quarks.

The $M_{t\bar{t}}$ distribution itself clearly remains the most powerful variable to search for evidence of resonant $t\bar{t}$ production; on the other hand, in case evidence for a $X \to t\bar{t}$ resonance is found in the much larger data sets to be collected in the near future at the LHC, the measurement of this polarization asymmetry as a function of $M_{t\bar{t}}$ could be used as well to corroborate such evidence and to possibly gain insight into the couplings of the new particle to the top quark.

6.1.4 Top quark $\cos\theta^*$ distribution

An important variable in the angular analysis of $t\bar{t}$ events is the so-called θ^* angle. The latter corresponds to the polar angle of the top quark momentum in the $t\bar{t}$ rest frame. This angular

distribution depends on the production mechanism of the $t\bar{t}$ system and it is expected to be sensitive to the spin of a hypothetical $t\bar{t}$ resonance [28]. Unlike the angular distributions discussed above, the $\cos\theta^*$ distribution is not related to the kinematics of the top quark decay products.

Figure 6.7 shows two independent sets of plots for the $\cos\theta^*$ distribution in data and MC. The plots on the left-hand side correspond to the expected and observed $\cos\theta^*$ distributions in the three categories of the ℓ + jets SR; in each category, the prediction for the SM background is found to be in good agreement with the values measured in data. The plots on the right-hand side contain, for each of the ℓ + jets SR categories, a comparison of the shape of the $\cos\theta^*$ distributions for SM $t\bar{t}$ production and a $\mathrm{Z}' \to t\bar{t}$ signal with $M_{\mathrm{Z}'} = 3$ TeV, considering only events with 2.5 TeV $< M_{t\bar{t}} <$ 3.5 TeV. This additional cut on $M_{t\bar{t}}$ would be naturally applied to study the properties of a newly discovered resonance (with $M_X \simeq 3$ TeV in this example). The plots show that, due to the correlation between the $\cos\theta^*$ variable and the invariant mass of the $t\bar{t}$ system, a cut on $M_{t\bar{t}}$ can strongly affect the distribution of this angular quantity, especially in the case of the SM $t\bar{t}$ background. In general, we find that the $\cos\theta^*$ distributions for background and signal exhibit a clear difference in shape and this difference can be further enhanced by applying a cut on $M_{t\bar{t}}$, compared to just using a more inclusive sample.

The $\cos\theta^*$ distribution is thus another angular observable that could be used to further reduce the background from SM $t\bar{t}$ production in the search for a $X \to t\bar{t}$ signal.

6.1.5 Impact of top quark helicity on expected sensitivity and projections at higher luminosities

In this section we make use of the MC signal samples introduced above to determine the expected limits on $\sigma(pp \to \mathrm{Z}') \cdot \mathrm{BR}(\mathrm{Z}' \to t\bar{t})$ for $\mathrm{Z}' \to t\bar{t}$ signals with different top quark polarizations, in the context of the ℓ + jets analysis described in Chapter 5. In addition, we estimate the projections of these expected sensitivities for higher values of the integrated luminosity, up to the $\mathscr{L} = 1000$ fb^{-1} foreseen for the High-Luminosity LHC, and we test one method to use top quark helicity observables to improve the sensitivity of the analysis.

Aside from considering different signal models, no changes are introduced in the statistical analysis with respect to the description given in Section 5.7, since the background model is not modified and nor is the treatment of systematic uncertainties. As a consequence, the cross section limits for signals with different top quark helicities reflect the difference in signal efficiency previously shown in Figure 6.2. Figure 6.8 shows the 95% CL expected limits on $\sigma \cdot \mathrm{BR}$ as a function of the Z' boson mass, for the three signal models (LH, RH and SM-like). As expected, the cross section limits for the RH case are improved by a factor between 20% and 35%, depending on the signal mass hypothesis, compared to the SM-like case; the limits for the LH case and the SM-like case, on the other hand, are found to be very similar to each other.

We consider now the projections of the expected sensitivity of the ℓ + jets analysis at higher values of the integrated luminosity. For these projections, differently to what is done in the standard analysis, we remove the uncertainty associated to the finite statistics of the MC samples, which is implemented in the statistical model via the simplified Barlow-Beeston method. In general, this effect is expected to be small when the MC samples have a limited statistical uncertainty compared to the data sets under consideration; this was indeed the case for the data analyzed in 2015, but future analyses with much higher integrated luminosity are expected to be performed with MC samples with much more statistics compared to the ones available in 2015. Thus, we remove this uncertainty, which is expected to be marginal,

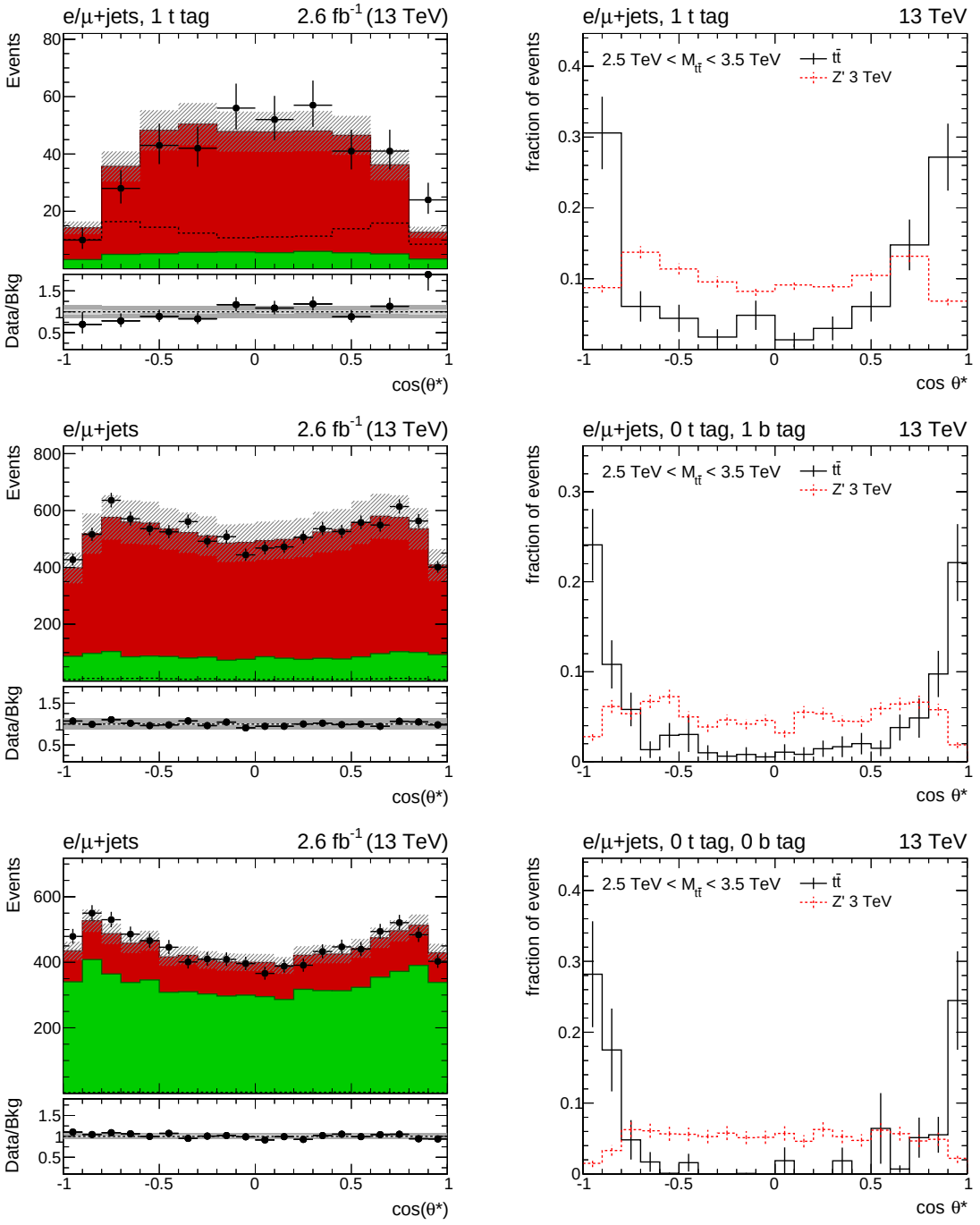


Figure 6.7: Reconstructed top quark $\cos\theta^*$ distributions in the ℓ + jets SR: (top) `T1` category, (center) `T0B1` category, (bottom) `T0B0` category. Each distribution includes the events of both the μ + jets and e + jets channels. Left: data/MC comparison for the $\cos\theta^*$ distribution of the reconstructed top quark in the ℓ + jets SR; the SM backgrounds are obtained from the likelihood fit described in Section 5.6. Right: shape comparison of the $\cos\theta^*$ distributions for SM $t\bar{t}$ production and a Z′ signal with $M_{Z'} = 3$ TeV, considering only events of the ℓ + jets SR with 2.5 TeV $< M_{t\bar{t}} <$ 3.5 TeV; each distribution in these plots is normalized to unity.

in order not to spoil these projections due to the limited size of the MC samples used in 2015.

Figure 6.9 shows the 95% CL expected limits on $\sigma \cdot \mathrm{BR}$ as a function of the projected integrated luminosity, for a $\mathrm{Z}' \to t\bar{t}$ resonance with SM-like couplings. The limits are calculated using, in one case, the same theoretical and experimental systematic uncertainties of the 2015 analysis, while, in the other case, removing all the systematic uncertainties in the statistical model. The first case can be considered pessimistic, as no improvement is assumed for any of the systematic errors; conversely, the case without systematic uncertainties corresponds to the best possible sensitivity that can be reached by the current analysis. These projections show that the expected limits for high-mass Z' signals ($M_{\mathrm{Z}'} \gtrsim 2$ TeV) are mainly limited by the statistical uncertainty, at least up to integrated luminosities around 300 fb^{-1}, where the impact of systematic errors on the limits becomes noticeable. Comparing these projections to the theoretical cross sections for a leptophobic $\mathrm{Z}' \to t\bar{t}$ signal in the topcolor model [139], these preliminary results suggest that approximately 200 fb^{-1} would be necessary to exclude a $\mathrm{Z}' \to t\bar{t}$ signal with $M_{\mathrm{Z}'} = 4$ TeV and a relative width of $\Gamma_{\mathrm{Z}'}/M_{\mathrm{Z}'} = 1\%$. Figure 6.10 shows the 95% CL expected limits for a $\mathrm{Z}' \to t\bar{t}$ signal (with SM-like couplings) as a function of the signal mass hypothesis, for a projected integrated luminosity of 1000 fb^{-1}. Since the SM-like, LH and RH cases differ only in the efficiency of the signal, at higher luminosities the relative difference between the exclusion limits in these three cases remains the same as the one shown in Figure 6.8.

As a possible way to improve the analysis sensitivity using information from the angular distributions of the top quark decay products, we consider splitting the events in the ℓ + jets SR by the sign of the $\cos\theta_\ell$ variable. This effectively doubles the number of $M_{t\bar{t}}$ spectra in the statistical analysis, as each SR category is divided in two samples, one with $\cos\theta_\ell > 0$ and the other with $\cos\theta_\ell < 0$. Figure 6.11 shows the relative variation in the cross section upper limits, for an integrated luminosity of 1000 fb^{-1}, between the standard event categorization and the new categorization in which the SR events are also split by the sign of $\cos\theta_\ell$, i.e. $(\sigma^{95\%\,\mathrm{CL}}_{\mathrm{split}} - \sigma^{95\%\,\mathrm{CL}}_{\mathrm{std}})/\sigma^{95\%\,\mathrm{CL}}_{\mathrm{std}}$.

The new categorization leads to an improvement between 10% and 20% for the exclusion limits in the LH case, while the limits in the RH case do not change significantly using this new configuration. This is due to the fact that, after all the analysis cuts are applied, the signal and background $\cos\theta_\ell$ distributions are characterized by fairly different shapes in the LH case, whereas in the RH case the two distributions are more similar to each other. Notably, this new categorization based on $\cos\theta_\ell$ provides a way to improve the expected limits for the LH case by roughly the same amount by which the expected limits in the RH case improve thanks to the higher signal efficiency of the RH signals.

Even though this simple test based on the use of the $\cos\theta_\ell$ variable leads only to a modest improvement in the expected limits, it serves as an example to show that angular variables can be exploited to improve the sensitivity of $X \to t\bar{t}$ searches. A larger improvement could possibly be achieved by combining the information from several angular distributions with more sophisticated techniques, e.g. multivariate analyses, in order to design an improved method to discriminate between signal and background.

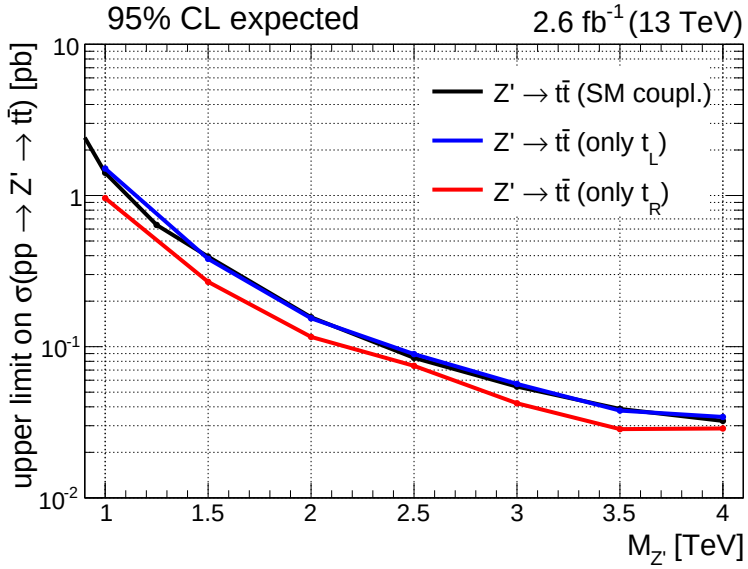


Figure 6.8: Expected limits on $\sigma(pp \to X) \cdot \mathrm{BR}(X \to t\bar{t})$ at 95% CL for $\mathrm{Z}' \to t\bar{t}$ signals with different top quark polarizations. The limits are calculated for an integrated luminosity of 2.6 fb^{-1}.

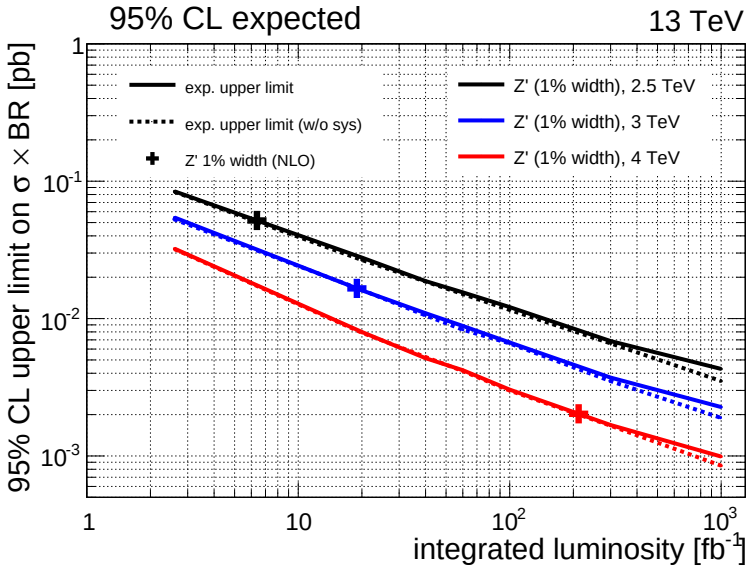


Figure 6.9: Expected limits on $\sigma(pp \to X) \cdot \mathrm{BR}(X \to t\bar{t})$ at 95% CL as a function of the projected integrated luminosity, for different signal mass hypotheses. The limits, obtained from the combination of the μ + jets and e + jets channels, refer to a $\mathrm{Z}' \to t\bar{t}$ signal with SM-like couplings to the top quark and a relative decay width of 1%. The solid (shaded) line correspond to the limit calculated with (without) systematic uncertainties. The cross symbol on each line indicates the corresponding theoretical cross sections for a leptophobic Z′ boson in the topcolor model [139].

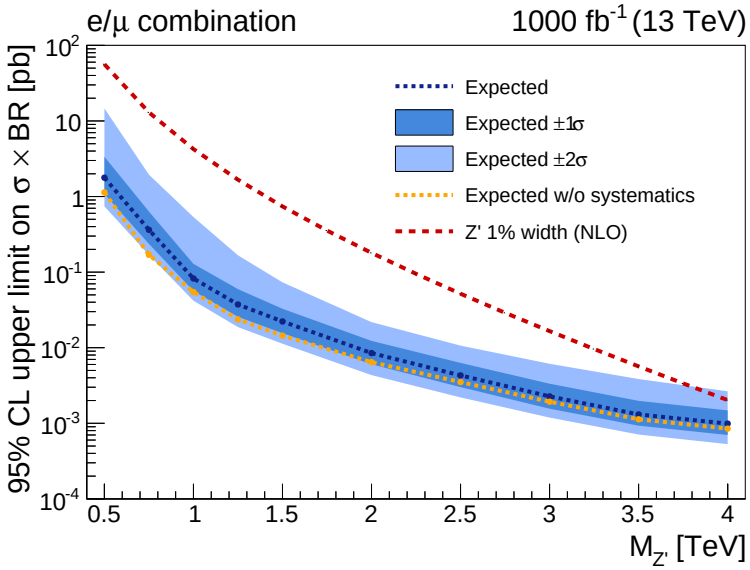


Figure 6.10: Expected upper limits on $\sigma(pp \to X) \cdot \mathrm{BR}(X \to t\bar{t})$ at 95% CL for a $\mathrm{Z}' \to t\bar{t}$ signal with SM-like couplings to the top quark and relative decay width of 1%. The limits, shown as a function of the signal mass hypothesis, are obtained from the combination of the $\mu+$jets and $e+$jets channels for a projected integrated luminosity of 1000 fb^{-1}. The theoretical predictions correspond to the cross sections for a leptophobic Z$'$ boson in the topcolor model [139].

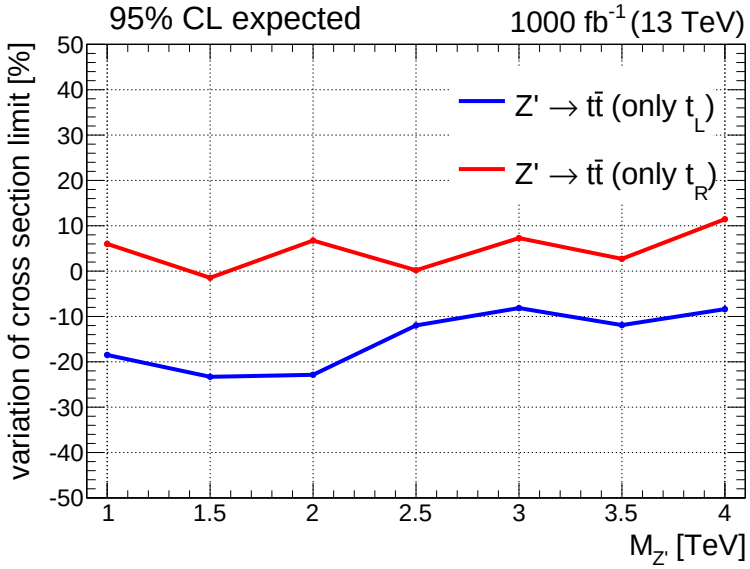


Figure 6.11: Variation in percentage of the expected limits on $\sigma(pp \to X) \cdot \mathrm{BR}(X \to t\bar{t})$, as a function of the signal mass hypothesis, after splitting the final event categories based on the sign of the $\cos\theta_\ell$ variable (compared to the standard categorization of the $\ell+$jets analysis), i.e. $(\sigma^{95\%\,\mathrm{CL}}_{\mathrm{split}} - \sigma^{95\%\,\mathrm{CL}}_{\mathrm{std}})/\sigma^{95\%\,\mathrm{CL}}_{\mathrm{std}}$.

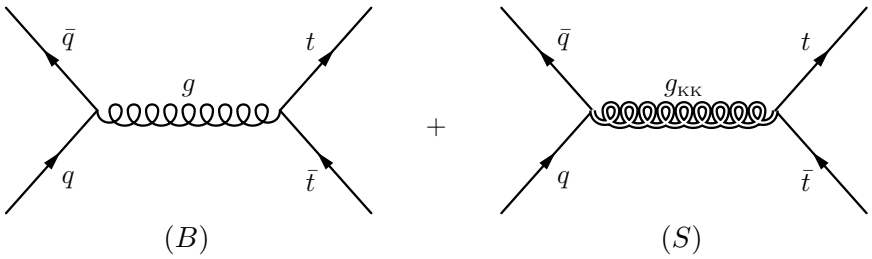


Figure 6.12: Tree-level diagrams contributing to the $q\bar{q} \to t\bar{t}$ process in a BSM scenario that includes a KK gluon resonance, g_{KK}.

6.2 Interference effects in the $M_{t\bar{t}}$ spectrum

The production of a generic $X \to t\bar{t}$ signal can in principle interfere with the production of a $t\bar{t}$ pair in the SM, leading to experimental signatures more complex than a simple resonant peak in the $M_{t\bar{t}}$ spectrum. In several BSM scenarios, these interference effects arise only at NLO and they are expected to be relatively small; the main example of such signals is that of a $\mathrm{Z}' \to t\bar{t}$ resonance, where the Z' is a color-singlet. On the other hand, there are other $X \to t\bar{t}$ BSM signals for which the interference term differs from zero already at tree-level. This is the case, for example, of a massive Higgs-like (spin-0) $t\bar{t}$ resonance and, also, of the color-octet KK gluon signal considered in the searches of Chapter 4 and Chapter 5. In this section we discuss the treatment of interference effects in $X \to t\bar{t}$ searches and we consider the KK gluon signal as a benchmark to quantify the impact of these effects on the results of the ℓ + jets analysis.

6.2.1 The KK gluon signal model

The MC samples used in this study for the production of a $g_{\mathrm{KK}} \to t\bar{t}$ signal in the RS model at $\sqrt{s} = 13$ TeV are produced with the PYTHIA-v8 event generator. In the following, we describe what are the main properties of this signal model and how predictions for its interference term with the SM $t\bar{t}$ background can be determined.

The tree-level diagrams for the production of a KK gluon decaying to a $t\bar{t}$ pair (S) and for $q\bar{q}$-initiated $t\bar{t}$ production in the SM (B) are shown in Figure 6.12. According to the implementation of the KK gluon model in PYTHIA-v8 [110], the cross section term coming from the interference of these two diagrams corresponds to

$$\sigma_{\mathrm{interf}}(q_i\bar{q}_i \to t\bar{t}) = \frac{16\pi\alpha_s^2}{27} \cdot \frac{g_V^i\, g_V^t\, A(\hat{s})\, \beta(\hat{s})\, \left(\hat{s} - M_{g_{\mathrm{KK}}}^2\right)}{\left(\hat{s} - M_{g_{\mathrm{KK}}}^2\right)^2 + \left(\hat{s}\frac{\Gamma_{\mathrm{tot}}}{M_{g_{\mathrm{KK}}}}\right)^2}\,, \tag{6.3}$$

where

$$A(\hat{s}) = 1 + 2m_t^2/\hat{s} \qquad \text{and} \qquad \beta(\hat{s}) = \sqrt{1 - 4m_t^2/\hat{s}}\,. \tag{6.4}$$

The index i in Equation (6.3) denotes the flavor of the initial-state quarks ($i = u, d, s, c, b$), $\hat{s}$ corresponds to the c.o.m. energy of the (parton-level) hard-scattering process, $M_{g_{\mathrm{KK}}}$ is the KK gluon mass hypothesis and Γ_{tot} denotes the KK gluon decay width (see Ref. [110] for its full expression). For each quark flavor q, the coefficient g_V^q corresponds to the vector coupling[1] of the KK gluon to the SM quark q, expressed in units of the QCD coupling g_s, with

[1]The vector and axial couplings of a SM quark are given by $g_V = (g_R + g_L)/2$ and $g_A = (g_R - g_L)/2$, respectively, where g_R (g_L) corresponds to its right-handed (left-handed) coupling.

$\alpha_s = g_s^2/4\pi$. As shown in Equation (6.3), the KK gluon interference term is proportional to the vector coupling of the KK gluon to the SM top quark.

The values of the KK gluon couplings to the SM quarks (in units of g_s) in the RS model correspond to

$$g_R^t = 5, \quad g_L^t = 1, \quad g_R^b = -1/5, \quad g_L^b = 1 \quad \text{and} \quad g_{R,L}^{u,d,c,s} = -1/5\,. \tag{6.5}$$

These are the couplings used in the MC simulation for the KK gluon signal. The most notable feature of this parameterization is that the g_R coupling of the KK gluon to the top quark is enhanced by a factor of five, whereas the g_L coupling is the same as that of the gluon in the SM. It is worth reminding that, even though the value of the KK gluon cross section depends on the choice of these couplings, the PYTHIA-v8 event generator does not take into account polarization effects in the decay of the final-state top quarks; as a consequence, the different proportion of left-handed and right-handed top quarks in the KK gluon decay is not reproduced correctly in the simulation. Based on the studies described in Section 6.1, the top quark polarization can affect significantly the sensitivity of the analysis; in particular, given the higher signal efficiency predicted for right-handed top quarks, the correct treatment of the top quark polarization would lead to a higher efficiency for a KK gluon signal in the RS model, compared to the results obtained for the $g_{\text{KK}} \to t\bar{t}$ signal samples in Ref. [4]. The same average fraction of left-handed and right-handed top quarks in the final state would be expected only for a KK gluon with pure vector couplings ($g_R = g_L$) to the top quark. In the case of pure axial couplings ($g_R = -g_L$), the interference term would be null before any event selection (due to its proportionality to g_V^t), but its net contribution after the cuts would differ from zero, due to the different efficiencies expected for right-handed and left-handed top quarks. Interference effects would be absent (at LO), before and after cuts, only if the KK gluon couplings to the initial-state quarks (u, d, s, c, b) were set to be pure axial couplings.

The KK gluon model in PYTHIA-v8 contains a functionality to simulate events with the full matrix element ($B + S$), thus giving the possibility to include interference effects. Exploiting this functionality, we perform a preliminary study to estimate the impact of the interference term in the KK gluon model on the results of the analysis. In order to do that, we adopt the following strategy:

1. we simulate parton-level events with PYTHIA-v8 considering, separately, the matrix elements for the KK gluon signal-only term ($|S|^2$), the $q\bar{q} \to t\bar{t}$ background term ($|B|^2$) and the full BSM model ($|B + S|^2$); from these simulations, the contribution of the interference term can be determined for any (generator-level) quantity as $I = |B + S|^2 - |S|^2 - |B|^2$;

2. we use the ratio of the interference term and the signal-only term to determine an interference-over-signal SF, for each KK gluon mass hypothesis, as a function of the invariant mass of the generated $t\bar{t}$ pair;

3. the interference-over-signal SFs are used to reweight the events in the (signal-only) $g_{\text{KK}} \to t\bar{t}$ MC samples and determine a prediction for the reconstructed $M_{t\bar{t}}$ distributions of the interference term in the ℓ + jets analysis.

In principle, the correct method to determine the interference term for the final reconstructed $M_{t\bar{t}}$ distributions in the ℓ+jets SR would be to produce the complete MC samples (up to the detector simulation) for the full BSM matrix element ($|B + S|^2$), the background-only term ($|B|^2$) and the signal-only term ($|S|^2$) and determine the interference term by subtraction

Table 6.1: Total cross sections predicted by PYTHIA-v8 for the full BSM model ($|B+S|^2$), the signal-only term ($|S|^2$) and the interference term ($|B+S|^2-|B|^2-|S|^2$). Each of these three values is shown for different KK gluon mass hypotheses. The total cross section predicted for the background-only term corresponds to 57.02 pb.

KK gluon cross sections [pb]

$M_{g_{\mathrm{KK}}}$ [GeV]	$\lvert B+S\rvert^2$	signal-only	interference
500	343.3	251.2	35.03
750	150.5	57.79	35.68
1000	97.04	18.69	21.33
1250	77.83	7.357	13.45
1500	69.37	3.324	9.021
2000	62.82	0.8723	4.926
2500	60.60	0.2903	3.288
3000	59.29	0.1150	2.157
3500	58.53	0.0518	1.458
4000	58.31	0.0260	1.266

using the reconstructed $M_{t\bar{t}}$ spectra obtained after the analysis selection. On the other hand, we can expect the strategy employed here (based on $M_{t\bar{t}}^{\mathrm{gen}}$ reweighting) to lead to sensible predictions, as long as the difference between the interference term and the signal-only term depends only on the mass of the generated $t\bar{t}$ system. This procedure also assumes that the efficiency of the analysis cuts for interference events and signal events with the same values of $M_{t\bar{t}}^{\mathrm{gen}}$ is on average the same; as long as this is the case, the interference-over-signal SFs do not depend on the analysis cuts.

6.2.2 Determination of the interference term at parton-level

The total cross sections predicted by PYTHIA-v8 for the signal-only simulation and the interference term are reported in Table 6.1 for different KK gluon mass hypotheses. Figure 6.13 shows the (parton-level) $M_{t\bar{t}}^{\mathrm{gen}}$ distributions, before any event selection, highlighting the difference between the KK gluon model with and without the inclusion of the interference term. The normalization of the interference term becomes larger and larger compared to the signal-only term, as the signal mass hypothesis increases; in particular, for the highest-mass signals ($M_{g_{\mathrm{KK}}} \gtrsim 3$ TeV), the interference term becomes the dominant part of the BSM prediction and it is characterized by a background-like shape which differs substantially from the signal-only distribution. As a consequence of the parameter choice given in Equation (6.5), the interference term is positive (constructive interference) for values of $M_{t\bar{t}}^{\mathrm{gen}}$ below the nominal g_{KK} mass, and negative (destructive interference) at higher mass values.

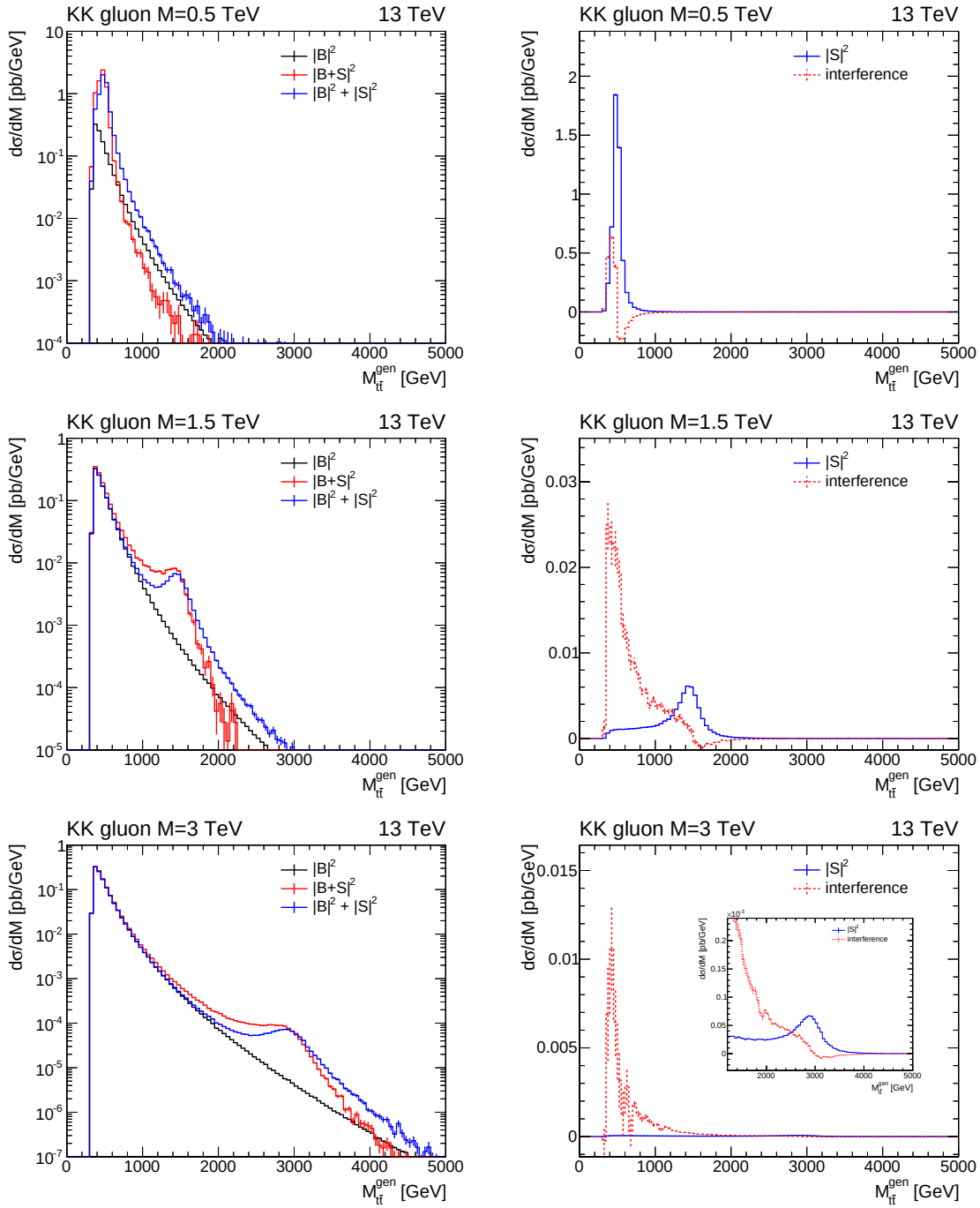


Figure 6.13: Differential distributions $d\sigma/dM_{t\bar{t}}^{\text{gen}}$, before any event selection, for different KK gluon mass hypotheses: (top) $M_{g_{\text{KK}}} = 0.5$ TeV, (center) $M_{g_{\text{KK}}} = 1.5$ TeV, (bottom) $M_{g_{\text{KK}}} = 3$ TeV. Each mass template is normalized to the corresponding cross section predicted by PYTHIA-v8. Left: comparison between the background-only model (black line), the full BSM model (red line) and the BSM model without interference effects (blue line). Right: comparison between the signal-only term $|S|^2$ (blue line) and the interference term, given by $|B+S|^2 - |B|^2 - |S|^2$ (shaded red line).

6.2.3 Impact of interference effects on the reconstructed $M_{t\bar{t}}$ distributions

The $M_{t\bar{t}}^{\text{gen}}$ distributions are used to determine the interference-over-signal correction factors. These values are then used to reweight the signal-only KK gluon MC as a function of $M_{t\bar{t}}^{\text{gen}}$ and estimate the expected contribution of the interference term in the ℓ + jets analysis. Figure 6.14 shows a comparison between the reconstructed $M_{t\bar{t}}$ distributions predicted for the KK gluon model, before and after including the interference term, as estimated with the procedure described above. For each mass hypothesis, these distributions are normalized to the cross sections of the KK gluon model determined with PYTHIA-v8, which correspond to $\mu_s = 1$. These plots show that, for all the mass hypotheses considered, the contribution of the interference term to the full BSM prediction is not negligible; in particular, as expected from the mass distributions obtained at generator-level, the interference term is the dominant component of the BSM distribution for the highest-mass signal hypotheses (for $\mu_s = 1$).

6.2.4 Exclusion limits on signal strength including interference effects and sensitivity projections at higher luminosities

The last part of this study focuses on the impact of interference effects on the final sensitivity of the analysis for the production of a KK gluon signal, considering first the case of the 2015 analysis and then the projections of the expected sensitivity for higher values of the integrated luminosity.

The statistical model defined in Section 5.7 needs to be extended in order to properly include the interference template in the limit calculation. For each bin used in the statistical analysis, the expected number of events corresponds to

$$N = b_0 + \mu_b \cdot b_1 + \mu_s \cdot s + \sqrt{\mu_b \cdot \mu_s} \cdot I \,, \tag{6.6}$$

where b_1 is the expected number of events for the $t\bar{t}$ background, b_0 corresponds to the expected yield of the sum of the other backgrounds, s denotes the signal-only prediction and I corresponds to the expectation for the interference term; μ_s is the standard signal strength modifier assigned to the signal-only distribution, while μ_b corresponds to the nuisance parameter associated to the normalization of the SM $t\bar{t}$ background. For simplicity, the fraction of $q\bar{q}$-initiated events in the SM $t\bar{t}$ background is fixed to prediction of PYTHIA-v8. Unlike the signal-only term, the interference term does not scale linearly with the signal strength modifier μ_s, but proportionally to $\sqrt{\mu_s}$; at the same time, the normalization of the interference term also scales proportionally to $\sqrt{\mu_b}$. It is worth noting that the expected yield for I, in a given bin, can be negative, but the overall expected yield N is always required to be higher than (or equal to) zero.

The THETA software package [131] has been modified in order to implement the proper treatment of the interference term in the statistical model. Since the introduction of the interference term spoils the direct proportionality between the expected number of signal events and the signal strength modifier μ_s, we present the results not in terms of upper limits on the signal cross section (like in Section 5.7), but in terms of upper limits on the μ_s parameter. We derive upper limits on the signal strength modifier μ_s at 95% CL using a Bayesian statistical method, similarly to the standard ℓ + jets analysis. The statistical procedure is based on a shape-analysis using the $M_{t\bar{t}}$ distributions of the six ℓ + jets SR categories. Systematic uncertainties for the background and signal predictions are included in the model according to the description given in Section 5.5 and Section 5.7. Similarly to the approach described in Section 6.1, these projections do not include the uncertainty associated to the limited statistics of the MC samples (Barlow-Beeston "light" method). One

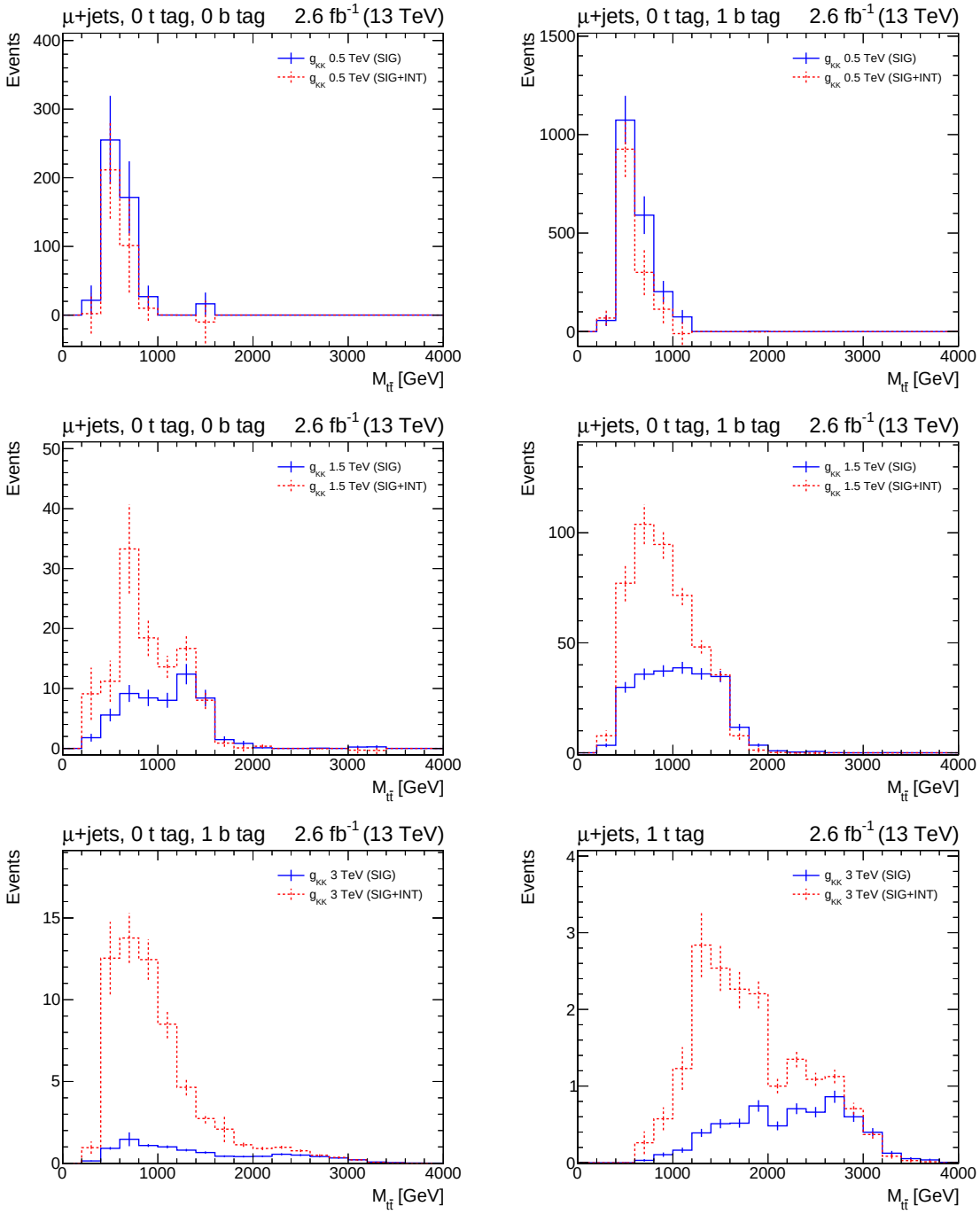


Figure 6.14: Reconstructed $M_{t\bar{t}}$ distributions predicted for the KK gluon model in selected categories of the μ + jets SR, before (solid blue line) and after (shaded red line) including the contribution of the interference term. Each distribution is normalized to the corresponding cross section predicted by the PYTHIA-v8 MC ($\mu_s = 1$). Plots are shown for different KK gluon mass hypotheses: (top) $M_{g_{\mathrm{KK}}} = 0.5$ TeV, (center) $M_{g_{\mathrm{KK}}} = 1.5$ TeV, (bottom) $M_{g_{\mathrm{KK}}} = 3$ TeV.

difference with respect to the strategy laid out in Section 5.7 is that the nuisance parameter associated to the $t\bar{t}$ cross section, which now also affects the normalization of the interference term, is left unconstrained. For the nominal prediction of the KK gluon model ($\mu_s = 1$), the cross sections obtained from PYTHIA-v8 are used for both the signal-only and interference templates.

Figure 6.15 shows the 95% CL upper limits on μ_s for the combination of the $\mu+$jets and $e+$jets channels, considering the signal-only term alone and the complete signal model including the interference term. For the case of the signal-only limits, the results correspond to the ones reported in Section 5.7, modulo the rescaling needed to transform the limit on μ_s into a limit on the absolute cross section of the process. It is worth noting that in this study we do not apply any K-factor to correct the LO signal cross sections, in order to make a direct comparison of the "signal-only" and the "signal+interference" cases without introducing any additional assumptions; as a consequence, the theoretical cross sections reported in Figure 5.50 correspond to the case $\mu_s = 1.3$ for the signal-only limits in Figure 6.15. For both cases ("signal-only" and "signal+interference") the observed limits on μ_s lie within the $\pm 2\sigma$ band associated to the expected limits. In the low-mass region ($M_{g_{\mathrm{KK}}} \lesssim 1.25$ TeV), the exclusion limits on μ_s do not change significantly after the introduction of the interference term: this is explained by the fact that, for these signal mass hypotheses, the contribution of constructive interference below the KK gluon mass is suppressed by the limited signal efficiency in the low-$M_{t\bar{t}}$ region; as a consequence, the addition of the interference term does not modify significantly the shape nor the total yield of the signal mass distributions. For the highest-mass signal hypotheses ($M_{g_{\mathrm{KK}}} \gtrsim 1.25$ TeV), the normalization of the interference term becomes increasingly large, compared to the signal-only prediction. In this case, the loss in the expected BSM yield due to destructive interference becomes a negligible effect and the addition of the interference term leads to an improvement in the final exclusion limits, even though the $M_{t\bar{t}}$ distributions of the interference component have a shape more similar to the background distributions.

We conclude this study by estimating how the introduction of the interference term can affect the exclusion limits expected for higher values of the integrated luminosity. Figure 6.16 shows the 95% CL expected limits on μ_s for a high-mass KK gluon signal, as a function of the integrated luminosity, before and after including the interference term. The plot shows that the correct treatment of interference effects leads to a clear improvement in the expected sensitivity. In particular, the relative improvement in the expected limits becomes bigger as the integrated luminosity increases and the expected limit on μ_s becomes smaller. This can be explained by the different dependence of the interference and signal-only terms on the value of the signal strength: in fact, for decreasing values of the signal strength ($\mu_s \ll 1$), the interference term scales only by a factor $\sqrt{\mu_s}$, whereas the signal-only term is reduced by a factor μ_s and becomes a subdominant contribution (despite the higher discriminating power of the signal-only mass distribution, compared to the background). Figure 6.17 shows the 95% CL expected limits on μ_s, for an integrated luminosity of 1000 fb^{-1}, comparing the results of the signal model with and without the interference term.

To summarize, this study shows that the introduction of the interference term for the KK gluon model is expected to have a significant impact on the final reconstructed $M_{t\bar{t}}$ distributions of the $\ell+$jets analysis. The corresponding effect on the final exclusion limits is not negligible, but, overall, modest for the integrated luminosity of the 2015 analysis. The impact of interference effects on the final sensitivity of the analysis are, in general, mitigated by the fact that the shape of the corresponding mass distributions is much closer to that of the SM background, compared to the signal-only spectra. Nevertheless, the introduction of

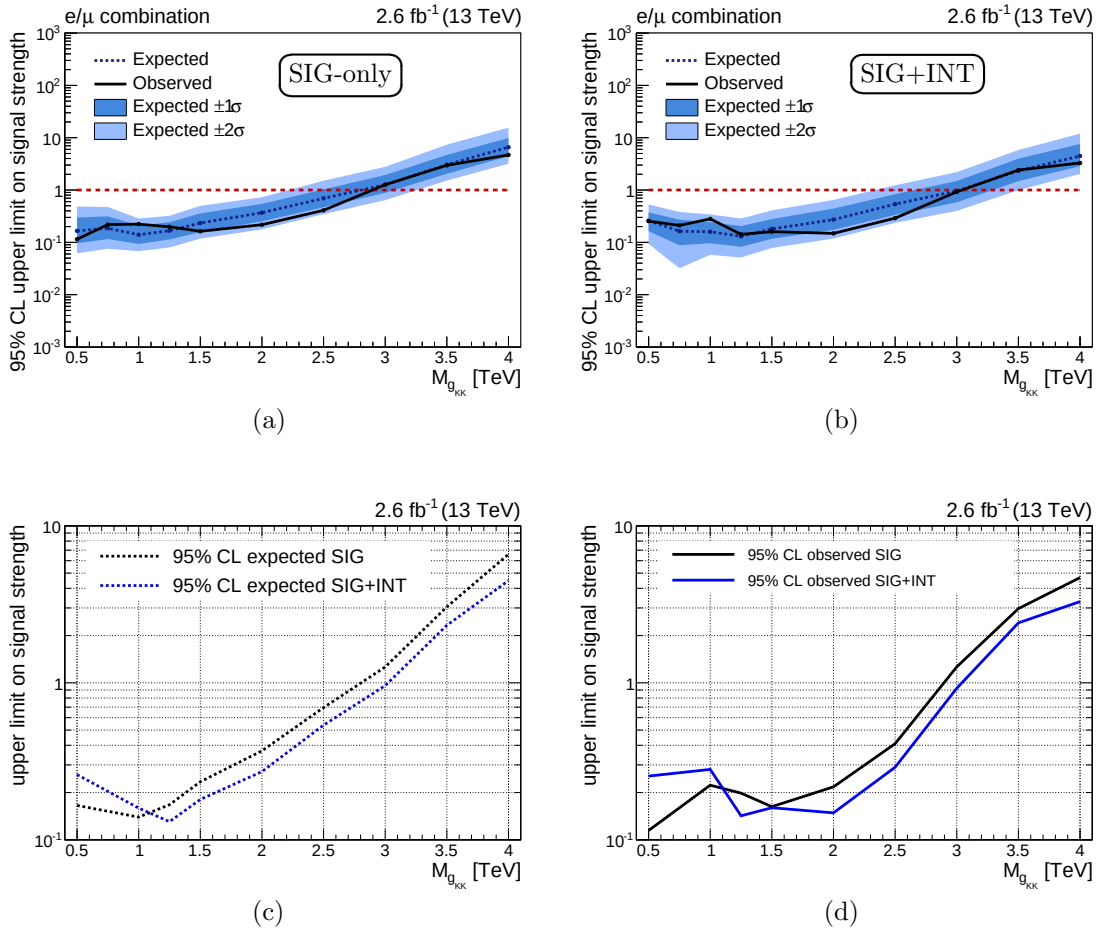


Figure 6.15: 95% CL Bayesian upper limits on the signal strength μ_s for the production of a KK gluon decaying to a $t\bar{t}$ pair: (a) KK gluon signal without interference effects, (b) full KK gluon model including the interference term, (c) comparison between the expected limits with and without the interference term, (d) comparison between the observed limits with and without the interference term. Limits are shown as a function of the KK gluon mass hypothesis.

interference effects is expected to lead to an improvement of the final limits in the high-mass region ($M_{g_{\mathrm{KK}}} \gtrsim 1.25$ TeV); based on this, the signal-only limits derived in Section 5.7 for the highest-mass signal hypotheses of a KK gluon resonance can be regarded as a conservative estimate. In addition, the correct treatment of interference effects is expected to have an increasingly large impact on the analysis sensitivity for higher values of the integrated luminosity, when exclusion limits will be set for values of the signal strength much below the theoretical prediction ($\mu_s < 1$).

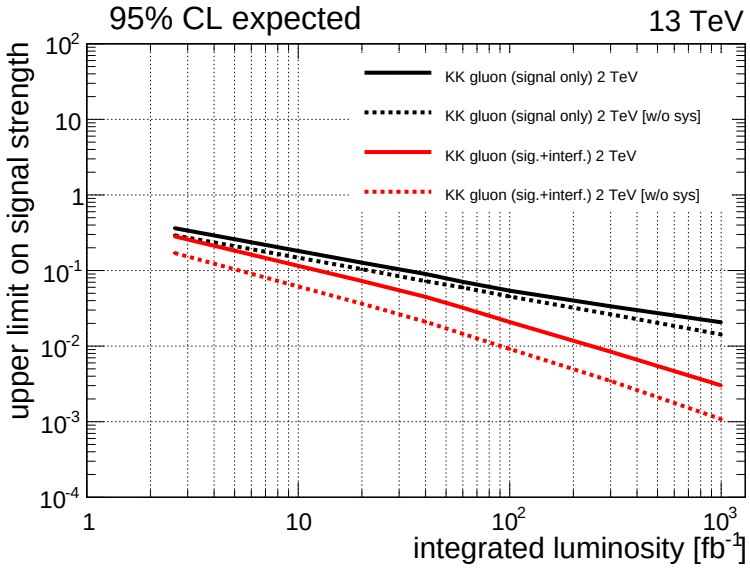


Figure 6.16: Comparison of the expected upper limits at 95% CL on the signal strength μ_s for a KK gluon signal with $M_{g_{\mathrm{KK}}} = 2$ TeV, with (red) and without (black) the interference term, as a function of the total integrated luminosity. The limits are determined from the combination of the $\mu+$jets and $e+$jets channels. The solid (shaded) line correspond to the limit calculated with (without) systematic uncertainties.

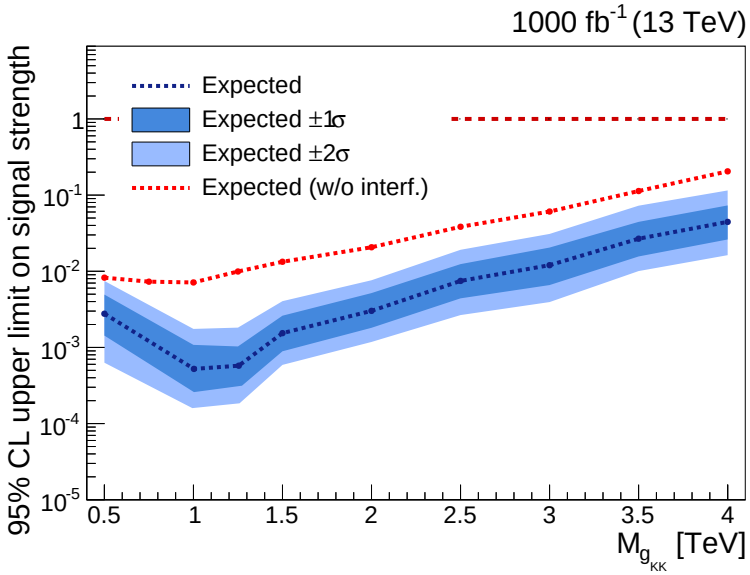


Figure 6.17: Expected upper limits at 95% CL on the signal strength μ_s for the KK gluon signal, after including the interference term, for an integrated luminosity of 1000 fb^{-1}. The limits, determined from the combination of the $\mu+$jets and $e+$jets channels, are shown as a function of the signal mass hypothesis. The dashed line corresponds to the expected limit for the KK gluon signal without interference effects.

Conclusions

This thesis has presented the results of two searches for new massive particles decaying to a $t\bar{t}$ pair with the CMS detector at the CERN LHC, using the $\ell(= e, \mu) + \text{jets}$ final state. Several new physics scenarios predict the existence of new particles that decay to a $t\bar{t}$ pair and appear in the $M_{t\bar{t}}$ spectrum as a resonant component on top of the distribution expected from SM processes. For this reason, the main objective of the two analyses is the experimental reconstruction of the $t\bar{t}$ system and the measurement of its invariant mass distribution.

The experimental searches described in this thesis show that there is no evidence for the production of a $X \to t\bar{t} \to \ell + \text{jets}$ signal in the data recorded by the CMS experiment in 2012 at $\sqrt{s} = 8$ TeV and in 2015 at the increased center-of-mass energy of $\sqrt{s} = 13$ TeV. The two searches thus set upper limits at 95% CL on the production cross section of a $X \to t\bar{t}$ resonance in a variety of BSM scenarios.

These results have been obtained by developing and testing state-of-the-art techniques for the reconstruction of top quarks produced in the LHC collisions with very large Lorentz boosts. These techniques include a dedicated method to identify non-isolated leptons from boosted top quark decays and the use of jet substructure algorithms to identify jets originating from hadronically decaying top quarks with high transverse momentum. This approach maximizes the analysis sensitivity for $t\bar{t}$ resonances with masses up to 4 TeV and, at the same time, it allows the study of $t\bar{t}$ pair production at very large values of the top quark momentum. In addition to these searches based on CMS data, possible future developments for $X \to t\bar{t}$ analyses have been investigated. The first of these studies focuses on the $t\bar{t}$ angular observables, showing how different top quark helicities can affect the signal efficiency of the $\ell + \text{jets}$ analysis and how the properties of a hypothetical $t\bar{t}$ resonance can be studied using angular distributions. A second study is concerned with the treatment of interference effects between a $t\bar{t}$ resonance and SM $t\bar{t}$ production, considering as an example a KK gluon resonance predicted in extradimensional models. This study shows that, for this specific BSM scenario, the correct treatment of interference effects leads to an improvement of the analysis sensitivity.

Despite the non-observation of a $X \to t\bar{t}$ signal in the data sets considered, the search for resonant $t\bar{t}$ production remains one of the most promising methods to discover new physics at the LHC. The analyses presented in this thesis significantly extend the exclusion limits set by similar searches performed in the past and they provide a solid baseline for future $X \to t\bar{t}$ searches, to be carried out in the coming years at the LHC.

Conclusiones

En esta tesis se han presentado los resultados de dos búsquedas de nuevas partículas masivas que se desintegran en un par $t\bar{t}$ con el detector CMS en el LHC del CERN, usando el estado final $\ell(= e, \mu)+$jets. Varios modelos de nueva física predicen nuevas partículas que se desintegran en un par $t\bar{t}$ y aparecen en el espectro $M_{t\bar{t}}$ como un componente resonante encima de la distribución esperada en los procesos del ME. Por esta razón, el objetivo principal de los dos análisis es la reconstrucción experimental del sistema $t\bar{t}$ y la medida de su distribución de masa invariante.

Las búsquedas experimentales descritas en esta tesis muestran que no hay pruebas de la producción de una señal de tipo $X \to t\bar{t} \to \ell +$ jets en los datos registrados por el experimento CMS en 2012 a $\sqrt{s} = 8$ TeV y en 2015 a la mayor energía en el centro de masas de $\sqrt{s} = 13$ TeV. Por lo tanto, las dos búsquedas ponen límites superiores a un nivel de confianza del 95% a la sección eficaz de producción de una resonancia $X \to t\bar{t}$ en el marco de varios modelos MME.

Estos resultados han sido obtenidos mediante el desarrollo y la aplicación de técnicas de vanguardia para la reconstrucción de quarks top de muy alto momento producidos en las colisiones del LHC. Estas técnicas incluyen un método para identificar leptones sin aislamiento provenientes de quarks top altamente energéticos y el uso de algoritmos capaces de analizar la estructura interna de los jets para identificar aquellos que se originan de la desintegración hadrónica de quarks top de alto momento transverso. Este enfoque maximiza la sensibilidad del análisis para resonancias $t\bar{t}$ con masas de hasta 4 TeV y, al mismo tiempo, permite estudiar la producción de pares $t\bar{t}$ con quarks top de elevado momento. Además de estas búsquedas basadas en los datos de CMS, se han investigado posibles futuros desarrollos en el análisis del canal $X \to t\bar{t}$. El primero de estos estudios se centra en los observables angulares del par $t\bar{t}$, mostrando como las diferentes helicidades del quark top pueden afectar la eficiencia del análisis y como las propiedades de una hipotética resonancia $t\bar{t}$ se pueden estudiar utilizando distribuciones angulares. Un segundo estudio se refiere al tratamiento de los efectos de interferencia entre una resonancia $t\bar{t}$ y la producción de un par $t\bar{t}$ en el ME, considerando como ejemplo el caso de un gluón KK masivo presente en modelos extradimensionales. Este estudio muestra que, para este modelo MME específico, el tratamiento correcto de los efectos de interferencia conlleva a una mejora en la sensibilidad del análisis.

A pesar de la ausencia de una señal de tipo $X \to t\bar{t}$ en las muestras de datos estudiadas, la búsqueda de producción resonante de un par $t\bar{t}$ sigue siendo uno de los métodos más prometedores para descubrir nueva física en el LHC. Los análisis presentados en esta tesis extienden considerablemente los límites de exclusión establecidos por búsquedas precedentes del mismo tipo y proporcionan una sólida base para futuras búsquedas en el canal $X \to t\bar{t}$, que se seguirán realizando en los próximos años en el LHC.

References

[1] **CMS** Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett. **B716** (2012) 30–61 [arXiv:1207.7235].

[2] **ATLAS** Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. **B716** (2012) 1–29 [arXiv:1207.7214].

[3] **CMS** Collaboration, "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV", Phys. Rev. **D93** (2016), no. 1 012001 [arXiv:1506.03062].

[4] **CMS** Collaboration, "Search for $\mathrm{t\bar{t}}$ resonances in boosted semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV", CMS-PAS-B2G-15-002 (2016).

[5] **Particle Data Group** Collaboration, "Review of Particle Physics", Chin. Phys. **C40** (2016), no. 10 100001.

[6] M. Peskin and D. Schroeder, "An Introduction To Quantum Field Theory", Frontiers in Physics Series, Westview Press (1995).

[7] D. Perkins, "Introduction to High Energy Physics", Cambridge University Press (2000).

[8] G. Dissertori et al., "Quantum Chromodynamics: High Energy Experiments and Theory", International series of monographs on physics, Clarendon Press (2003).

[9] K. G. Wilson, "The Renormalization Group and Strong Interactions", Phys. Rev. **D3** (1971) 1818.

[10] G. 't Hooft, "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking", NATO Sci. Ser. B **59** (1980) 135–157.

[11] G. F. Giudice, "Naturalness after LHC8", PoS **EPS-HEP2013** (2013) 163 [arXiv:1307.7879].

[12] G. Bertone et al., "Particle Dark Matter: Evidence, Candidates and Constraints", Phys. Rept. **405** (2005) 279–390 [hep-ph/0404175].

[13] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces", Phys. Rev. Lett. **32** (1974) 438–441.

[14] R. W. Robinett and J. L. Rosner, "Minimally extended electroweak gauge theories in SO(10) and E6", AIP Conf. Proc. **99** (1983) 193–201.

[15] D. London and J. L. Rosner, "Extra Gauge Bosons in E(6)", Phys. Rev. **D34** (1986) 1530.

[16] S. Dimopoulos and L. Susskind, "Mass Without Scalars", Nucl. Phys. **B155** (1979) 237–252.

[17] E. Eichten and K. D. Lane, "Dynamical Breaking of Weak Interaction Symmetries", Phys. Lett. **B90** (1980) 125–130.

[18] A. Leike, "The Phenomenology of extra neutral gauge bosons", Phys. Rept. **317** (1999) 143–250 [hep-ph/9805494].

[19] P. Langacker, "The Physics of Heavy Z' Gauge Bosons", Rev. Mod. Phys. **81** (2009) 1199–1228 [arXiv:0801.1345].

[20] C. T. Hill, "Topcolor: Top quark condensation in a gauge extension of the standard model", Phys. Lett. **B266** (1991) 419–424.

[21] C. T. Hill, "Topcolor assisted technicolor", Phys. Lett. **B345** (1995) 483–489 [hep-ph/9411426].

[22] L. Randall and R. Sundrum, "A large mass hierarchy from a small extra dimension", Phys. Rev. Lett. **83** (1999) 3370–3373 [hep-ph/9905221].

[23] L. Randall and R. Sundrum, "An alternative to compactification", Phys. Rev. Lett. **83** (1999) 4690–4693 [hep-th/9906064].

[24] K. Agashe et al., "LHC Signals from Warped Extra Dimensions", Phys. Rev. **D77** (2008) 015003 [hep-ph/0612015].

[25] T. D. Lee, "A Theory of Spontaneous T Violation", Phys. Rev. **D8** (1973) 1226–1239.

[26] D. Dicus et al., "Higgs decay to top quarks at hadron colliders", Phys. Lett. **B333** (1994) 126–131 [hep-ph/9404359].

[27] **ATLAS** Collaboration, "Search for heavy Higgs bosons A/H decaying to a top-quark pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", ATLAS-CONF-2016-073 (2016).

[28] R. Frederix and F. Maltoni, "Top pair invariant mass distribution: A Window on new physics", JHEP **01** (2009) 047 [arXiv:0712.2355].

[29] **CDF** Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", Phys. Rev. Lett. **74** (1995) 2626–2631 [hep-ex/9503002].

[30] **D0** Collaboration, "Observation of the top quark", Phys. Rev. Lett. **74** (1995) 2632–2637 [hep-ex/9503003].

[31] **ATLAS**, **CDF**, **CMS** and **D0** Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass", ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416 (2014) [arXiv:1403.4427].

[32] N. Kidonakis, "NNLL threshold resummation for top-pair and single-top production", Phys. Part. Nucl. **45** (2014), no. 4 714–722 [arXiv:1210.7813].

[33] N. Kidonakis, "Top Quark Production", Proceedings of "Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)" (2014) 139–168 [arXiv:1311.0283].

[34] D. E. Kaplan et al., "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks", Phys. Rev. Lett. **101** (2008) 142001 [arXiv:0806.0848].

[35] J. Thaler and L.-T. Wang, "Strategies to Identify Boosted Tops", JHEP **07** (2008) 092 [arXiv:0806.0023].

[36] **CDF** Collaboration, "Search for Resonant Top-Antitop Production in the Lepton Plus Jets Decay Mode Using the Full CDF Data Set", Phys. Rev. Lett. **110** (2013), no. 12 121802 [arXiv:1211.5363].

[37] **D0** Collaboration, "Search for a Narrow $t\bar{t}$ Resonance in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. **D85** (2012) 051101 [arXiv:1111.1271].

[38] **ATLAS** Collaboration, "Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV", Phys. Rev. **D88** (2013), no. 1 012004 [arXiv:1305.2756].

[39] **CMS** Collaboration, "Search for resonant $t\bar{t}$ production in lepton+jets events in pp collisions at $\sqrt{s} = 7$ TeV", JHEP **12** (2012) 015 [arXiv:1209.4397].

[40] **CMS** Collaboration, "The CMS experiment at the CERN LHC", JINST **3** (2008) S08004.

[41] O. S. Brüning et al., "LHC Design Report", CERN-2004-003-V-1 (2004).

[42] L. Evans and P. Bryant, "LHC Machine", JINST **3** (2008) S08001.

[43] **CMS** Collaboration, "CMS Physics: Detector Performance and Software", Technical Design Report (Volume 1), CERN-LHCC-2006-001 (2006).

[44] **CMS** Collaboration, "CMS Physics: Physics Performance", Technical Design Report (Volume 2), CERN-LHCC-2006-021 (2007).

[45] R. Assmann et al., "A brief history of the LEP collider", Nucl. Phys. Proc. Suppl. **109B** (2002) 17–31.

[46] **ALICE** Collaboration, "The ALICE experiment at the CERN LHC", JINST **3** (2008) S08002.

[47] N. Cabibbo and G. Parisi, "Exponential Hadronic Spectrum and Quark Liberation", Phys. Lett. **B59** (1975) 67–69.

[48] **LHCb** Collaboration, "The LHCb Detector at the LHC", JINST **3** (2008) S08005.

[49] **ATLAS** Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", JINST **3** (2008) S08003.

[50] **CMS** Collaboration, "The CMS tracker system project", Technical Design Report, CERN-LHCC-98-006 (1997).

[51] **CMS** Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", JINST **9** (2014), no. 10 P10009 [arXiv:1405.6569].

[52] **CMS** Collaboration, "The CMS electromagnetic calorimeter project", Technical Design Report, CERN-LHCC-97-033 (1997).

[53] **CMS** Collaboration, "The CMS hadron calorimeter project", Technical Design Report, CERN-LHCC-97-031 (1997).

[54] **CMS** Collaboration, "Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data", JINST **5** (2010) T03012 [arXiv:0911.4991].

[55] **CMS** Collaboration, "The CMS muon project", Technical Design Report, CERN-LHCC-97-032 (1997).

[56] **CMS** Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC", JINST **8** (2013) P11002 [arXiv:1306.6905].

[57] **CMS** Collaboration, "The TriDAS project: The Trigger Systems", Technical Design Report (Volume 1), CERN-LHCC-2000-038 (2000).

[58] **CMS** Collaboration, "The TriDAS Project: Data Acquisition and High-Level Trigger", Technical Design Report (Volume 2), CERN-LHCC-2002-026 (2002).

[59] J. Ero et al., "The CMS Drift Tube Trigger Track Finder", JINST **3** (2008) P08006.

[60] **CMS Trigger and Data Acquisition** Group, "The CMS high level trigger", Eur. Phys. J. **C46** (2006) 605–667 [hep-ex/0512077].

[61] **CMS** Collaboration, "The CMS trigger system", JINST **12** (2017), no. 01 P01020 [arXiv:1609.02366].

[62] **CMS** Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update", CMS-PAS-SMP-12-008 (2012).

[63] S. van der Meer, "Calibration of the Effective Beam Height in the ISR", CERN-ISR-PO-68-31 (1968).

[64] **CMS** Collaboration, "The Computing project", Technical Design Report, CERN-LHCC-2005-023 (2005).

[65] C. Eck et al., "LHC computing Grid", Technical Design Report, CERN-LHCC-2005-024 (2005).

[66] **CMS** Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001 (2009).

[67] **CMS** Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", CMS-PAS-PFT-10-001 (2010).

[68] **CMS** Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from *pp* Collisions at 7 TeV", CMS-PAS-PFT-10-002 (2010).

[69] **CMS** Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", CMS-PAS-PFT-10-003 (2010).

[70] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", Nucl. Instrum. Meth. **A262** (1987) 444–450.

[71] **CMS** Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS-PAS-TRK-10-005 (2010).

[72] K. Rose, "Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems", Proceedings of the IEEE **86** (1998), no. 11 2210–2239.

[73] T. Speer et al., "Vertex Fitting in the CMS Tracker", CMS-NOTE-2006-032 (2006).

[74] R. Frühwirth et al., "Adaptive Vertex Fitting", CMS-NOTE-2007-008 (2007).

[75] **CMS** Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", JINST **7** (2012) P10002 [arXiv:1206.4071].

[76] **CMS** Collaboration, "Search for ttH production in multilepton final states at $\sqrt{s} = 13$ TeV", CMS-PAS-HIG-15-008 (2016).

[77] **CMS** Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", JINST **10** (2015), no. 06 P06005 [arXiv:1502.02701].

[78] **CMS** Collaboration, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", CMS-NOTE-2005-001 (2005).

[79] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", Phys. Rev. **D48** (1993) 3160–3166 [hep-ph/9305266].

[80] M. Cacciari et al., "The anti-k_t jet clustering algorithm", JHEP **04** (2008) 063 [arXiv:0802.1189].

[81] **CMS** Collaboration, "Pileup Removal Algorithms", CMS-PAS-JME-14-001 (2014).

[82] **CMS** Collaboration, "Jet energy scale and resolution in the CMS experiment in *pp* collisions at 8 TeV", Submitted to JINST (2016) [arXiv:1607.03663].

[83] **CMS** Collaboration, "Performance of the CMS missing transverse momentum reconstruction in *pp* data at $\sqrt{s} = 8$ TeV", JINST **10** (2015), no. 02 P02006 [arXiv:1411.0511].

[84] **CMS** Collaboration, "Identification of b-quark jets with the CMS experiment", JINST **8** (2013) P04013 [arXiv:1211.4462].

[85] **CMS** Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events", CMS-PAS-BTV-13-001 (2013).

[86] **CMS** Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS-PAS-BTV-15-001 (2016).

[87] **CMS** Collaboration, "Boosted Top Jet Tagging at CMS", CMS-PAS-JME-13-007 (2014).

[88] **CMS** Collaboration, "Top Tagging with New Approaches", CMS-PAS-JME-15-002 (2016).

[89] **CMS** Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging", CMS-PAS-JME-09-001 (2009).

[90] A. J. Larkoski et al., "Soft Drop", JHEP **05** (2014) 146 [arXiv:1402.2657].

[91] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", JHEP **03** (2011) 015 [arXiv:1011.2268].

[92] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", JHEP **02** (2012) 093 [arXiv:1108.2701].

[93] T. Sjostrand et al., "PYTHIA 6.4 Physics and Manual", JHEP **05** (2006) 026 [hep-ph/0603175].

[94] **CMS** Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV", JHEP **09** (2011) 109 [arXiv:1107.0330].

[95] **GEANT4** Collaboration, "GEANT4: A Simulation toolkit", Nucl. Instrum. Meth. **A506** (2003) 250–303.

[96] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", JHEP **11** (2004) 040 [hep-ph/0409146].

[97] S. Frixione et al., "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", JHEP **11** (2007) 070 [arXiv:0709.2092].

[98] S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP **06** (2010) 043 [arXiv:1002.2581].

[99] M. Czakon et al., "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$", Phys. Rev. Lett. **110** (2013) 252004 [arXiv:1303.6254].

[100] J. Alwall et al., "MadGraph 5 : Going Beyond", JHEP **06** (2011) 128 [arXiv:1106.0522].

[101] M. L. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", JHEP **01** (2007) 013 [hep-ph/0611129].

[102] R. Gavin et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", Comput. Phys. Commun. **182** (2011) 2388–2403 [arXiv:1011.3540].

[103] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", Nucl. Phys. Proc. Suppl. **205-206** (2010) 10–15 [arXiv:1007.3492].

[104] H.-L. Lai et al., "New parton distributions for collider physics", Phys. Rev. **D82** (2010) 074024 [arXiv:1007.2241].

[105] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", JHEP **07** (2002) 012 [hep-ph/0201195].

[106] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", Phys. Rev. **D78** (2008) 013004 [arXiv:0802.0007].

[107] B. Lillie et al., "The Bulk RS KK-gluon at the LHC", JHEP **09** (2007) 074 [hep-ph/0701166].

[108] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation", JHEP **09** (2007) 028 [arXiv:0706.2334].

[109] T. Sjostrand et al., "A Brief Introduction to PYTHIA 8.1", Comput.Phys.Commun. **178** (2008) 852–867 [arXiv:0710.3820].

[110] S. Ask et al., "Identifying the colour of TeV-scale resonances", JHEP **01** (2012) 018 [arXiv:1108.2396].

[111] M. Cacciari et al., "A note on the fate of the Landau–Yang theorem in non-Abelian gauge theories", Phys. Lett. **B753** (2016) 476–481 [arXiv:1509.07853].

[112] R. M. Harris and S. Jain, "Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop", Eur. Phys. J. **C72** (2012) 2072 [arXiv:1112.4928].

[113] J. Gao et al., "Next-to-leading order QCD corrections to the heavy resonance production and decay into top quark pair at the LHC", Phys. Rev. **D82** (2010) 014020 [arXiv:1004.0876].

[114] **CMS** Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", JINST **6** (2011) P11002 [arXiv:1107.4277].

[115] **CMS** Collaboration, "Jet Energy Corrections determination at 7 TeV", CMS-PAS-JME-10-010 (2010).

[116] **CMS** Collaboration, "Jet Energy Resolution in CMS at sqrt(s)=7 TeV", CMS-PAS-JME-10-014 (2011).

[117] **CMS** Collaboration, "Search for heavy narrow resonances decaying to $t\bar{t}$ in the muon+jets channel", CMS-PAS-EXO-11-055 (2011).

[118] **CMS** Collaboration, "Search for high-mass resonances decaying to $t\bar{t}$ in the electrons+jets channel", CMS-PAS-EXO-11-092 (2012).

[119] **CMS** Collaboration, "Search for $t\bar{t}$ resonances in semileptonic final states in pp collisions at $\sqrt{s} = 8$ TeV", CMS-PAS-B2G-12-006 (2013).

[120] **CMS** Collaboration, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s} = 8$ TeV", Phys. Rev. Lett. **111** (2013), no. 21 211804 [arXiv:1309.2030]. [Erratum: Phys. Rev. Lett. **112** (2014), no. 11 119903].

[121] **CMS** Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in *pp* collisions at 8 TeV", CMS-PAS-TOP-12-027 (2013).

[122] **CMS** Collaboration, "Measurement of the differential ttbar cross section in the dilepton channel at 8 TeV", CMS-PAS-TOP-12-028 (2013).

[123] **CMS** Collaboration, "Differential cross section measurements for the production of a W boson in association with jets in proton–proton collisions at $\sqrt{s} = 7$ TeV", Phys. Lett. **B741** (2015) 12–37 [arXiv:1406.7533].

[124] **CMS** Collaboration, "Measurement of the production cross section for a W boson and two b jets in *pp* collisions at $\sqrt{s} = 7$ TeV", Phys. Lett. **B735** (2014) 204–225 [arXiv:1312.6608].

[125] **CMS** Collaboration, "Observation of the associated production of a single top quark and a W boson in *pp* collisions at sqrt(s) = 8 TeV", Phys. Rev. Lett. **112** (2014) 231802 [arXiv:1401.2942].

[126] **CMS** Collaboration, "Measurement of the differential production cross section of Z bosons in association with jets in *pp* collisions at sqrt(s) = 8 TeV", CMS-PAS-SMP-13-007 (2014).

[127] **CMS** Collaboration, "Measurement of the W^+W^- and ZZ production cross sections in *pp* collisions at $\sqrt{s} = 8$ TeV", Phys. Lett. **B721** (2013) 190–211 [arXiv:1301.4698].

[128] **CMS** Collaboration, "Measurement of WZ and ZZ production in *pp* collisions at $\sqrt{s} = 8$ TeV in final states with b-tagged jets", Eur. Phys. J. **C74** (2014), no. 8 2973 [arXiv:1403.3047].

[129] **CMS** Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS-PAS-LUM-13-001 (2013).

[130] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", J. Phys. **G43** (2016) 023001 [arXiv:1510.03865].

[131] J. Ott, The statistical framework and documentation can be found at http://www.theta-framework.org.

[132] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", Comput. Phys. Commun. **77** (1993) 219–228.

[133] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", Proceedings of "PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding" (2011) [arXiv:1103.0354].

[134] T. Plehn et al., "Stop Reconstruction with Tagged Tops", JHEP **10** (2010) 078 [arXiv:1006.2833].

[135] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", Comput. Phys. Commun. **191** (2015) 159–177 [arXiv:1410.3012].

[136] **CMS** Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", Eur. Phys. J. **C76** (2016), no. 3 155 [arXiv:1512.00815].

[137] **NNPDF** Collaboration, "Parton distributions for the LHC Run II", JHEP **04** (2015) 040 [arXiv:1410.8849].

[138] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", JHEP **07** (2014) 079 [arXiv:1405.0301].

[139] R. Bonciani et al., "Electroweak top-quark pair production at the LHC with Z' bosons to NLO QCD in POWHEG", JHEP **02** (2016) 141 [arXiv:1511.08185].

[140] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", Comput. Phys. Commun. **185** (2014) 2930 [arXiv:1112.5675].

[141] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", Comput. Phys. Commun. **191** (2015) 74–89 [arXiv:1406.4403].

[142] T. Gehrmann et al., "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD", Phys. Rev. Lett. **113** (2014), no. 21 212001 [arXiv:1408.5243].

[143] **CMS** Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", CMS-PAS-FSQ-15-005 (2016).

[144] **CMS** Collaboration, "Search for a Narrow Resonance Produced in 13 TeV pp Collisions Decaying to Electron Pair or Muon Pair Final States", CMS-PAS-EXO-15-005 (2015).

[145] **CMS** Collaboration, "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV", Phys. Lett. **B713** (2012) 68–90 [arXiv:1202.4083].

[146] K. Rehermann and B. Tweedie, "Efficient Identification of Boosted Semileptonic Top Quarks at the LHC", JHEP **1103** (2011) 059 [arXiv:1007.2221].

[147] **CMS** Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", CMS-PAS-LUM-15-001 (2016).

[148] G. Mahlon and S. J. Parke, "Angular correlations in top quark pair production and decay at hadron colliders", Phys. Rev. **D53** (1996) 4886–4896 [hep-ph/9512264].